Head Rotation Detection in Marmoset Monkeys

by

Sravanthi Simhadri

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2014 by the
Graduate Supervisory Committee:

Pavan Turaga, Co-Chair
Yi Zhou, Co-Chair
Visar Berisha

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Head movement is known to have the benefit of improving the accuracy of sound localization for humans and animals. Marmoset is a small bodied New World monkey species and it has become an emerging model for studying the auditory functions. This thesis aims to detect the horizontal and vertical rotation of head movement in marmoset monkeys.

Experiments were conducted in a sound-attenuated acoustic chamber. Head movement of marmoset monkey was studied under various auditory and visual stimulation conditions. With increasing complexity, these conditions are (1) idle, (2) sound-alone, (3) sound and visual signals, and (4) alert signal by opening and closing of the chamber door. All of these conditions were tested with either house light on or off. Infra-red camera with a frame rate of 90 Hz was used to capture of the head movement of monkeys. To assist the signal detection, two circular markers were attached to the top of monkey head. The data analysis used an image-based marker detection scheme. Images were processed using the Computation Vision Toolbox in Matlab. The markers and their positions were detected using blob detection techniques. Based on the frame-by-frame information of marker positions, the angular position, velocity and acceleration were extracted in horizontal and vertical planes. Adaptive Otsu Thresholding, Kalman filtering and bound setting for marker properties were used to overcome a number of challenges encountered during this analysis, such as finding image segmentation threshold, continuously tracking markers during large head movement, and false alarm detection.

The results show that the blob detection method together with Kalman filtering yielded better performances than other image based techniques like optical flow and SURF features .The median of the maximal head turn in the horizontal plane was in the range of 20 to 70 degrees and the median of the maximal velocity in horizontal plane was in the range of a few hundreds of degrees per second. In comparison, the natural alert signal – door opening and closing – evoked the faster head turns than other stimulus conditions. These results suggest that behaviorally relevant stimulus such as alert signals evoke faster head-turn responses in marmoset monkeys.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Head movement is a natural behavior of humans and animals in response to a sound stimulus and in an attempt to localize it. For many animals, sound localization is a survival skill. The ability to quickly identify the location of an approaching predator or an escaping prey improves the fitness of a species over the course of evolution. The obvious benefit of head movement is improved accuracy of sound localization. Moving the head towards a sound source is an excellent strategy in that spatial hearing yields higher acuity and discriminability when the source is located directly in front than those in lateral and rear positions (Blauert 1997; Middlebrooks and Green 1990). This is especially important for listening in situations where the sound source is out of sight. Furthermore, head movement helps improve sound analyses by minimizing font-back confusions (Wallach 1940; Wightman and Kristler 1997). For humans and many animal species, the left or right direction of a sound wave is determined by the timing and level differences between sound waves reaching the two ears, while the elevation angle of a sound wave is determined by the spectral filtering of the outer ear. However, localization errors often occur because the binaural disparity cues, interaural time and level differences, are not uniquely related to sound locations (Wightman and Kristler 1997). Front and back locations that reside on the same sagittal plane could yield identical binaural cue. As such, a listener cannot distinguish a sound source that is located in front from the other that is located in the back, which is commonly known as the "cone of confusion". Head movement can remove such ambiguity by breaking the balance and enlarging the difference in the binaural cues of the two sources.

While extensive research has been conducted to understand the spatial hearing performances of humans and animals and the neural basis of sound localization, we know little about the way in which the brain processes spatial information of sounds during head movement and the related physiological mechanisms for localization improvement discussed above.

To further research in this direction, this thesis will examine the head movement of the common marmoset (*Callithrix jacchus*). Marmoset is a small-bodied New World monkey species. Marmosets rely on hearing to communicate and navigate. In their natural habitats, marmosets live in small groups (20, *Behavior of the common marmoset, nd*) They communicate with each other using visual cues (i.e. moving body parts) or vocal calls (Epple 1968). There are many situations in which they depend on their auditory system, such as communicating location of colony, warning about predators, during play. Marmosets are diurnal and spend most of the day looking for food. Because of these natural conditions the marmoset monkeys are in general very aware of their surroundings.

An interesting theory for the speed of locomotion in animals is given by Spoor *et al.* (2007). According to them, the body movements depend on the size of the semicircular canal of the inner ear. This study was performed on different mammal species including primates. It was observed that animals with large canals compared to body mass index, made quick jerky movements. Based on measurement taken by Johnson *et al.* (2012), marmosets are high energy species with relatively large semi-circular canal, and make rapid movement.

In recent years, marmoset became an emerging model for studying auditory and visual cortical functions of non-human primates (Bendor & Wang, 2005; Burman *et al.*, 2010; Cloherty *et al.*, 2010). Unlike many rodent animal models, marmoset monkey has similar brain organization and structural anatomy to humans. Additionally, the marmoset monkey has acute hearing with a hearing range that largely overlaps with the humans (Osmanski and Wang 2011; Seiden 1957) and it tends to make rapid head turns in response to auditory and visual stimuli even in a captive environment. Previous studies have shown that marmosets appear to use both binaural and spectral cues to localize a sound (Slee and Young 2010), and the spatial selectivity of neurons is highly restricted in the marmoset auditory cortex (Zhou and Wang, 2010). These features make marmoset an ideal model for studying spatial hearing during head movement.

The objective of this research is to measure and characterize the kinetics of head movements of marmoset monkeys. The method chosen is centered on image-based motion detection, and specifically marker-based detection. This project used marker-based tracking of images captured by an infrared camera. A few other methods were explored and ruled out as inappropriate for the requirements of this project. The research procedures are designed to overcome multiple challenges that are unique to this project. (1) The speed and dynamic range of head motions that marmoset monkeys make exceed those of humans. Existing techniques used for humans (e.g., feature based detection and tracking algorithms) are not adequate for the present research. (2) The head size of the marmoset is very small, about 5 cm in diameter. This feature puts restrictions on the setup, such as the distance of camera to the monkey's head, size of the markers that can be used. (3) Marmosets without head restraining devices are chosen to study the natural

motions. As a result, special considerations have to be taken for the types of markers, weights of markers and numbers of markers that can be used. (4) Marmoset monkeys used here are the ones with limited behavioral training in order to study the natural reflexive responses to the stimulus presented. Since an animal is not required to respond, the experiments need to collect a large amount of data to obtain meaningful comparisons among results taken from different testing conditions.

There are many methods of measuring the head rotation angles and speed of rotation as summarized in Chapter 2. Chapter 3 gives a detailed explanation of handling of the animals and the experimental setup. Chapters 4-5 report the results of the research. The three variables that were investigated are (1) horizontal and vertical rotation angles, and the corresponding (2) angular velocity and (3) acceleration. The head movements were separately analyzed in the horizontal and vertical planes and result analyses focus on the horizontal movement. The outcome of this research may provide critical information about head turning behavior of the marmoset monkey species for both neurophysiological and behavioral research.

CHAPTER 2: BACKGROUND

2.1 Head Motion

Figure 1 shows the coordinate system used for sound localization experiments. The horizontal plane is the plane through the head and parallel to the ground, the frontal plane is parallel to the face and perpendicular to the ground and the median plane cuts through the nose and perpendicular to the face and ground.
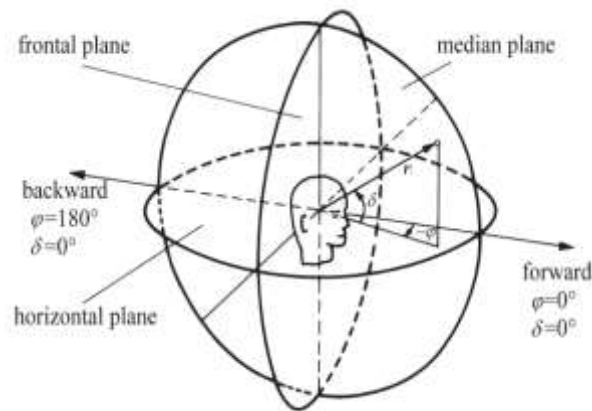
Figure 1: Head Coordinate System Used in Auditory Spatial Experiments as Shown in Blauert, J. (1997) and in the Current Thesis.

A previous research on head movements for localization was performed by Thurlow et al (1967), where they tried to study the nature of head motions in humans in response to auditory stimulus. They classify different types of head movements and also study the reversal of head movements. They used a camera-based system and the analysis was subjective determination of motion from the video. This was an early system and now with advancements in science and technology there are more robust systems.

2.2 Current Methods and Products

Since the objective of this thesis is to study the horizontal and vertical rotations of the monkey's head and the corresponding velocities and accelerations, one of the logical approaches is to first consider the monkey head as a general solid object and explore the various methods. In other words, the head orientation problem is reduced to an object tracking problem. The other approach is of course, is to detect the head and determine the orientation.

2.2.1 Methods

Existing technologies used a number of methods to perform object tracking. Below is a summary of basic approach and the advantages and disadvantages of each approach:

- **Directional Suggestion**: This method is mentioned in the survey report by Murphy-Chutorian & Trivedi (2009) requires asking a subject to move his/her head towards a marker in a room. Since the location of the marker is known, the 3-dimensional direction /angle where the person is looking can be obtained easily. An improved version is the directional suggestion with laser pointer mounted on the subjects' head. The advantage of the system is that it is simple to construct. However, this is a poor method in terms of accuracy and it is extremely difficult to do with animals.

- **Manual Annotation**: This method is also given in the mentioned in the survey report by Murphy-Chutorian & Trivedi (2009), is often used to establish ground truth (which in the case of object tracking is a dataset showing the exact orientation information for a subset of the data and becomes the tool for

6

evaluating algorithms). Images of faces are viewed by a person and are assigned labels of different orientations. This method cannot be used for fine pose estimation because it would be cumbersome.

- **Camera Arrays**: This method uses multiple cameras to capture images of the object from different angles. For the orientation estimation scenario this method requires the object to be present in a fixed location in the area under the camera array. The camera array can be bundled with a software package and such a system is called motion capture system which is a robust multi-camera tracking system. Advantages of this system are that it gives real-time data and 6 DoF rotation parameters. The disadvantages are it would require marker suits to be worn by subject which is not suitable for a monkey.

- **Magnetic sensors**: This method uses emitting and detecting magnetic fields to obtain orientation information. Advantages of such a system are that it is affordable and has a high accuracy of less than 1 degree. The disadvantages are that signals from sensors are susceptible to noise and can be interfered by the presence of metals.

  Zangemeister *et al*. (1981) studied the head rotations for gaze and eye-head coordination and used similar sensors. Guitton and Volle (1987) in their study on gaze control measured head and eye movements in humans. For the purpose of head movement detection, they made use of a search coil which is a magnetic field sensor.

- **Inertial Sensors**: This method uses accelerometers and gyroscopes to determine orientation. The principle behind the accelerometer is it detects the changes in

linear acceleration and tilt angle. A widely used application of inertial sensors is in smart phones to reorient the screen with respect to changes in viewing angles of a user. The gyroscope works on the principle of measuring the angular rate of rotation. Sometimes a certain number of accelerometers gyroscopes are combined into one system.

Advantages of this method are that it is not very expensive and the devices are small since the marmoset monkey is small, it would be an advantage to have small sensors. Also it gives 6 DoF data in real-time. Disadvantages of this method are that it cannot be mounted on a naïve monkey, and would require head restraining system on the monkey's head like a head implant.

- **Appearance template method**: This method, mentioned in the survey report by Murphy-Chutorian & Trivedi (2009), compares new image to set of labelled templates (the labels being orientations) and selects most similar among the templates to get current orientation. Advantages of this method are that they do not require negative training (as in feature detection) and are suitable for images with either low or high resolution. Disadvantages of this method is that interpolation is required which adds to computation. The inherent assumption is that head is detected; if there is no detection, the method fails. Also if the number of templates is large, computation increases.

- **Geometric approaches:** These methods are also introduced in the survey report by Murphy-Chutorian & Trivedi (2009) and make use of head shape and orientation can be estimated from ratios of facial points such as outside corners of nose, outside corners of mouth and length of nose. Yaw angle is determined by

8

difference of these ratios. Pitch angle is obtained by comparing distance between nose tip and eye-line to an anthropometric model. Advantage of this method is that it is fast and simple. Disadvantage is that it is not suitable for near frontal views and also precision is not very good.

2.2.2 Commercial Motion Capture Systems

There are a number of commercially available motion capture systems. The following are only a few of the different systems.

- **NaturalPoint TrackIR** (17**,** *Natural point TrackIR*, nd**)**

  The primary application of this system is in gaming. It uses an infrared camera which tracks the head to change the view of the user in a game simulation. The system uses reflective markers attached to the cap of the user. It needs the user to be at a distance of 24-36 inches from the camera. It has the advantages and disadvantages of a motion capture system.

- **Visage|SDK™ HeadTrack** (26, *Visage technologies tracking and animation, nd*)

  The system gives real-time 3D translation and rotation and facial feature points at the rate of 30 frames per second and includes gaze determination. The system is fully automated, robust to occlusions, and does not require markers. The underlying technology is based on fitting a 3D model to the facial image and estimating the 3D motion of the head and the facial expression. The same model will not be applicable for marmoset monkey head tracking.

- **IGS-190™ Motion Capture System by MetaMotion** (13, *IGS-190 motion capture system, nd*)

This system uses the accelerometer-gyro setup described in the previous section. It has a fast learning curve of 5 minutes and requires little calibration and hence is easy to use. It makes use of 18 'gyro' sensors and measures rotations of person's bones in real-time. This system has the same disadvantages as discussed for magnetic sensors.

- **WII Nintendo system**

 The primary application of this system is gaming. The core components are the Wii Remote controller (28, *What is wii U?, nd*) and the Wii console. The Wii Remote hosts the motion sensors connected to the console using Bluetooth technology and the range for which it can be used is 10 meters.

It is the system used for motion tracking in the research of Brimijoin *et al* (2013), where the authors make use of a custom infrared LED array for the purpose of tracking. The LED Array is placed on the head of the human listening subject and the remote is placed about 1.5m above the head. This method is also not suitable here because of the equipment cannot be placed on the monkey's head.

- **Kinect**

The Kinect comes with a Microsoft Windows SDK with a face tracking application. It uses both the depth sensor and the color camera. The human face model used is called the Candide model and certain features of the face are matched to that of the model. The advantages of this device are that it gives a robust face tracking. Disadvantages of using Kinect are it cannot track marmoset monkey face without training and training would be a cumbersome process. The

dynamic range of the face tracking is limited to +/- 40 degrees in yaw and roll and +/- 20 degrees in pitch. Thus using a single Kinect would not be sufficient. Also the frame rate of Kinect is around 30 fps which is not sufficient for detecting the fast head turns that the marmoset makes.

2.3 Software Techniques

All the above methods are the general object detection and tracking methods. Some are specific to head detection. Following are the techniques which are general computer vision algorithms implemented in this research and some background study that is essential in understanding and implementing them.

2.3.1 Blob detection

The core idea behind marker detection and tracking is blob detection. Blob is a binary large object that represents a region of interest with certain properties compared to surrounding regions. One type of blob detection makes use of Laplacian of Gaussian which is the convolution of image with a Gaussian kernel and applying the Laplacian operator. The size of the Gaussian kernel determines the size of the blob that can be detected. To be able to detect different sized blobs scale normalized Laplacian operator can be used. The technique used is connected components in binary images. Connectivity can be 4 or 8 as shown in figure 2. Grass-fire algorithm is used to form connected regions.
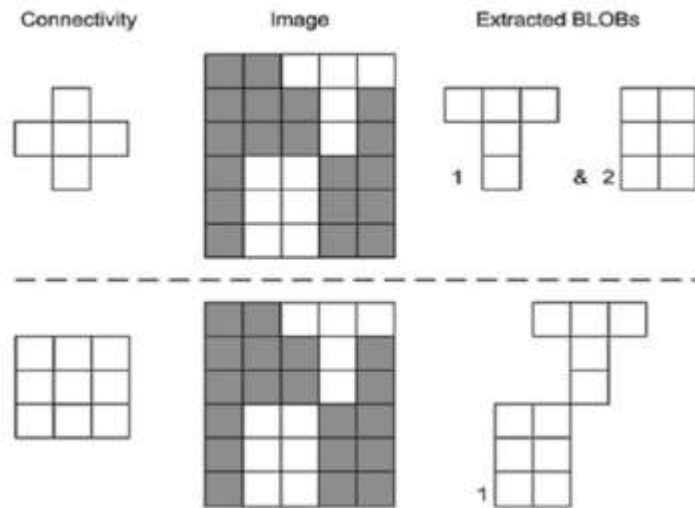
Figure 2: Blob Connectivity.

Blob connectivity in a part of an image (what-when-how 2014). Top image shows 4-connectivity where two objects are detected and bottom image shows 8- connectivity where one object is detected.

2.3.3 Optical Flow

Optical flow estimation is determining the 2D motion fields from time varying intensities in image. The techniques used are based on gradient estimation as in Fleet & Weiss (2005), block matching and correlation.

Optical flow uses the Taylor series expansion of image signals, making use of partial derivatives with respect to time and space coordinates. The aperture can be overcome by the Lucas Kanade method. The assumption made is that flow is constant in the neighborhood of pixels. Another method used to solve the constraint problem is the Horn Schunck method- It is a global method, assumes smoothness in flow over entire image. It is a dense optical flow solution since flow is determined for every pixel. Because of this, it becomes computationally expensive.

12

The assumptions for Lucas-Kanade algorithm taken from Bradsky & Kaehler (2008)

1. Brightness constancy- There is no change in intensities of image pixels.

2. Movements are small- to be able to track the motion, object movement must not be large.

3. Spatial coherence- neighboring points on an image correspond to objects in scene with similar motion.

Gradient estimation:

Consider a point (x,y) in an Image I which undergoes a translation in the x and y directions by u and v respectively and the result is H (Seitz, 2009).

$$H(x,y) = I (x+u, y+v)$$

The expression is expanded using the Taylor Series, ignoring higher order terms.

$$I(x+u,y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order differentials}$$

$$I(x+u,y+v) \cong I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

$$\text{Therefore, } H(x,y) \cong I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

$$0 \cong I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v - H(x,y)$$

$$0 \cong I(x,y) + I_x u + I_y v - H(x,y)$$

Note that I_x is the partial derivative with the x-direction, I_y is the partial derivative with the y-direction and I_t is the partial derivative with time t. The above equation can be denoted by matrix representation. It can be seen that there are more variables in the resulting equation and it cannot be solved as it is.

$$0 \cong I_t + I_x u + I_y v \cong I_t + \nabla I .[u \ v] \tag{2}$$

13

To be able to solve equation (2), we make use of the brightness constancy where we can get more equations with the same variables. Thus making use of assumptions in a window of given size the equations can be rewritten as:

$$\begin{bmatrix} I_x(p1) & I_y(p1) \\ \vdots & \vdots \\ I_x(p16) & I_y(p16) \end{bmatrix} * \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p1) \\ \vdots \\ I_t(p16) \end{bmatrix}$$

Assuming window size is four, there are now 16 equations for solving the two unknowns. These equations will be solvable if there are corners in the image to track, to avoid singular matrix.

2.3.4 Geometric Transforms

Transforms are used for making alterations to images and objects inside images in terms of translation, rotation, shear, etc. In the current scenario, these transforms help in identifying the horizontal/vertical head rotation based on results of other algorithms like optical flow.

Following are the different types of transformations (11,*Geometric Transformations, nd*):

1. Euclidean Transformation

   The transformation could be a translation, rotation or reflection. This transformation preserves lengths and angles measured as well as the shapes of objects and can be used to detect a change in position and orientation of objects.
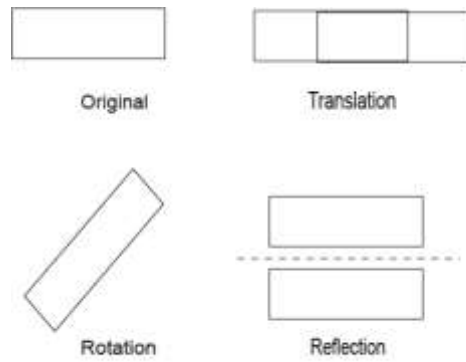
Figure 3: Euclidean Transforms in Translation, Rotation and Reflection.
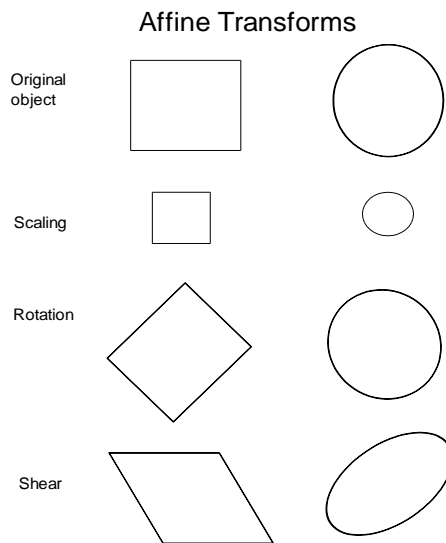
Figure 4: Affine Transforms in Scaling, Rotation and Shear

2. Affine transformation

This is a generalization of Euclidean transformation. Apart from above transformations it includes scaling and shear. Here lengths and angles are not preserved and the degree of a polynomial is not altered, parallel lines/planes are

transformed to parallel lines/planes, and intersecting lines/planes are transformed
to intersecting lines and planes. Shape of the geometric object will change.

3. Projective Transforms

They could be considered as generalization of affine transforms. They require
homogeneous co-ordinate system. If (x,y,z,w) is the homogeneous coordinate of a
3D point, where w = 1, then the 3D point is given by (x/w,y/w,z/w,1) Projective
transformation can bring points at infinity to a finite range and finite points to
infinity.

Point matching is the method by which given two images, interest points are matched
between them to find the transformation that was applied.

2.3.5 SURF

It is the acronym for Speeded Up Robust Features. It is based on SIFT- Scale Invariant
Feature Transform and is faster than SIFT. The important aspect of this method is the
interest point detector. One example of such a detector is the Harris corner detector which
is not scale invariant. Other examples of interest points are blobs and T-junctions. A good
interest point should be repeatable i.e. it must continue to be an interest point in different
conditions of illumination, viewing directions, and noise as proposed by *Bay et al(2006).*
A scale invariant version is given by Lowe using Laplacian of Gaussian approximated by
Difference of Gaussian.

SIFT descriptor is a distribution of smaller scale features of neighborhood of interest
point. The method is robust to deformations. Computes histogram of local oriented
gradients around the interest point and stores the bins. SURF approximates Laplacian of
Gaussian with box filter. Convolution with box filter is fast when using integral images

and can be done in parallel for different scales. For rotation invariance wavelet responses are used. The neighborhood of every interest point is stored as a feature vector.

For matching the feature vectors, we make use of either Euclidean or Mahalanobis distances. For this research, Euclidean distance is used. Speed of this algorithm depends on dimensionality of feature vector. But reducing the dimension drastically will affect the distinctiveness of the features.

SURF focuses on scale invariance and rotation invariance and skew. Here interest points are selected based on Hessian calculation which is a square matrix of second order partial derivatives of the intensity. The Hessian matrix for the pixel $\dot{x}$(x,y) for scale σ is written in terms of Laplacian function as

$$H(\dot{x}, \sigma) = \begin{bmatrix} L_{xx}(\dot{x}, \sigma) & L_{xy}(\dot{x}, \sigma) \\ L_{xy}(\dot{x}, \sigma) & L_{yy}(\dot{x}, \sigma) \end{bmatrix}$$

where $L_{xx}(\dot{x}, \sigma)$ is the convolution of Gaussian second order derivative with image at $\dot{x}$.

To implement rotation invariance, Haar wavelet features are made use of. The wavelet responses are computed for different scale values which turn out to be the criterion for feature matching. Here also integral images are used for speed.

2.3.6 Kalman Filtering

In object tracking, there would be scenarios where the system would have to estimate the location of the object. Also the accuracy of the estimate would be affected by accuracy of sensor, occlusions, and these can be considered as noise. There is a motion model associated with the object being tracked, like, constant velocity, constant acceleration.

There are three broad categories of motion to be considered for Kalman filter tracking according to Bradsky & Kaehler (2008),

- Dynamic motion- motion depends on previous "state". Object location in the next state is given by velocity times the current location.

- Control motion- the user has some control over the motion, by application of a known external force. An example is in the area of robotics (tracking the location of robot) where a control signal is given to the robot.

- Random motion- because of this motion, the variance of the next state estimate increases. Here again the assumption is that the random motion is Gaussian.

The state of the system can be modeled in a number of ways. For tracking, the most used parameters are position, velocity, and acceleration. The important underlying assumption for any of the above categories is that the system being modeled is linear. The second important assumption is that the measurement noise is white Gaussian noise.

Figure 5 shows the two phases of working of the filter taken from Bradsky & Kaehler (2008): prediction and correction. In the prediction phase, an estimate is made about the object location based on the existing measurement data. In the correction phase, the predicted value is updated with incoming measurements. The information of system at time $t+dt$ is fused with information of system at time t projected at time $t+dt$. Kalman filter can also be used to give the best estimate by combining measurements from different sources.
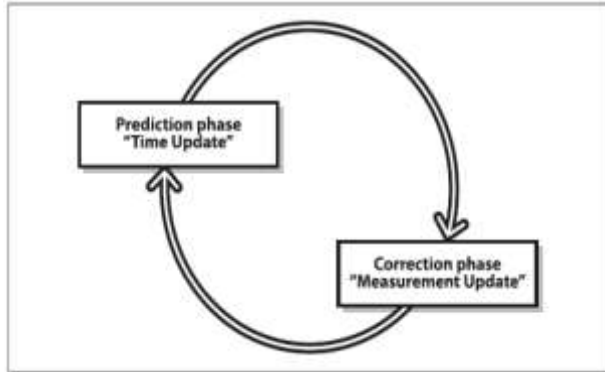
Figure 5: Basic Principle of Kalman Filter as Shown in Chapter 10, Bradsky & Kaehler (2008).

This method assumes instead of keeping the history of measurements, an algorithm that can iteratively update the state and utilize this updated state for next step can be written. After a measurement is made, the model parameters are updated. The system should take into account the uncertainty of the existing model and the uncertainty of the incoming measurements and make a prediction so as to maximize the accuracy of the predicted measurements.

CHAPTER 3: EXPERIMENTAL PROCEDURES

3.1 Animal Preparation and Positioning

Experiments were conducted in awake common marmoset monkeys (*Callithrix jacchus*).
Experimental procedures were approved by the Institutional Animal Care and Use
Committee of the Arizona State University following NIH (National Institutes of Health)
guidelines. Some of the chief handling procedures are summarized here. These steps are
important for animal safety as well as the safety of experimenters.

All tests were conducted in a sound chamber that provides 60 dB noise attenuation and is
equipped with audio-visual instrumentation. The experimenter wears Personal Protective
Equipment (PPE) at all time while conducting the experiment. Animals that participated
in the tests were all trained to sit in the custom-made primate chair and to be handled by
an experimenter based on detailed training procedures developed in the lab. Animals
were rewarded with juice for maintaining desired behaviors in chair.

An animal was brought to the sound chamber using a transport carrier. It takes 2 weeks
for an animal to get acclimated to the transport carrier. While moving the monkey from
the transport carrier into the chair, the experimenter carefully grabs the animal by
wrapping a gloved hand around the waist of the animal and guides it into the chair cavity.
Prior to conducting experiments, the monkey is adapted to sit calmly in the chair. It is
important to keep track of the animals being use for experiment. The behavior of each
animal may be different, and especially for those naïve, un-trained monkeys. All
experiment steps and animal behavioral and health conditions were documented in the lab
book.

The marmoset is a small bodied new world monkey with an average height of 18 cm and correspondingly a small head. Table 1 shows the list of monkeys used in the experiment and the basic information of head size and body weight. This is important information for guiding the usage of markers for head tracking.

Table 1: List of Lab Monkeys Used for this Experiment and Their Basic Information.

Monkey Name	Gender	Average weight (in gm)	Head (ear-ear) (in cm)	Head (top-chin) (in cm)	Head (back-front fat-pad) (in cm)
Eeyore	Male	457	4.3	3.5	5.2
Dot	Female	417	3.5	3.0	4.0
Akela	Female	256	3.2	3.0	4.2
Bumblebee	Male	366	3.0	3.2	4.0

Before performing the experiment, the details of each monkey such as the name and weight are recorded.

3.2 Experimental Setup

Figure 6 shows the experimental setup. Experiments were conducted in a dimly lit, double-walled, sound-attenuated chamber (IAC, Industrial Acoustics, 7'x7'x6.6'). The internal walls and ceiling were lined with ~3 inch acoustic absorption foam (Sonex, Illbruck). Two Loudspeakers (Adam F5), which have a relatively flat spectrum between 50 – 50k Hz, were positioned at left and right angles $\pm45°$ relative to the center of a subject head. The acoustic signals were programmed in custom-made programs written in Matlab (Mathworks) and converted to analog signals and sent to the loudspeakers via an AD/DA device USB Pre. Images are sent from a low-noise Pico projector to an acoustically transparent screen (33x55 in cm^2, Seymour AV). The viewing distance was 45 cm with 0.78 cm on the screen for one visual degree.

21
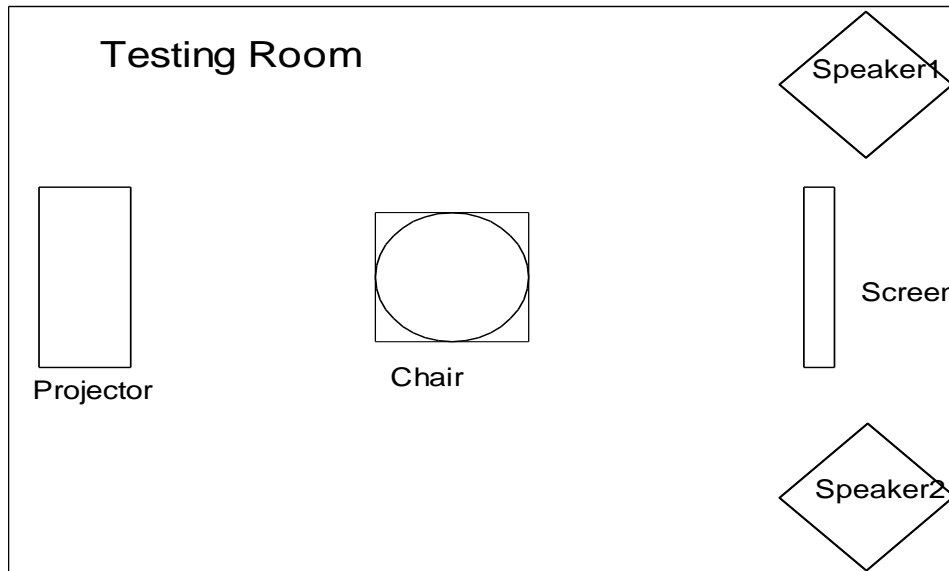
Figure 6: Schematic Drawing of the Experimental Setup and a Snap-shot of the Interior Layout of the Acoustic Chamber.

Sound equipment was calibrated using a sound level meter (Brüel & Kjær Type 2250) during the setup to deliver sounds from the computer control programs to loudspeakers with expected sound pressure levels (SPLs). When the calibration was complete, the volume controls in equipment including USB pre and loudspeakers were disabled. The

SPL of a stimulus was solely controlled by the computer program. The calibration procedures were repeated for pure tones at frequencies between 1 and 8 kHz and for broadband noise.

The marmoset monkey sat in the primate chair in the center of the acoustic chamber as seen in Figure 7. Although the body was constrained, the animal could move its head in any directions at will. Head position of a subject was detected by an infrared eye-tracking system (Arrington Research) with a frame rate of 90 Hz and an image resolution of 480 by 640 pixels. The infra-red camera was placed directly above the animal at a height of 15 cm. The illuminator (light source) was placed next to the camera to illuminate the field of view of the camera. The field of view was chosen so that the head with the markers could be captured in each frame. However this was not always accomplished once the animal tilted its head to extreme lateral positions. Strategies to compensate for situations like this are discussed in later sections. The raw video files were stored in AVI format and analyzed in Matlab using the Computer Vision toolbox for video specific development.

The usual order of experiment was to start the video recording, then start the stimulus presentation. After a certain number of runs of stimulus (on an average of ten), stop the stimulus presentation and finally stop the video recording.
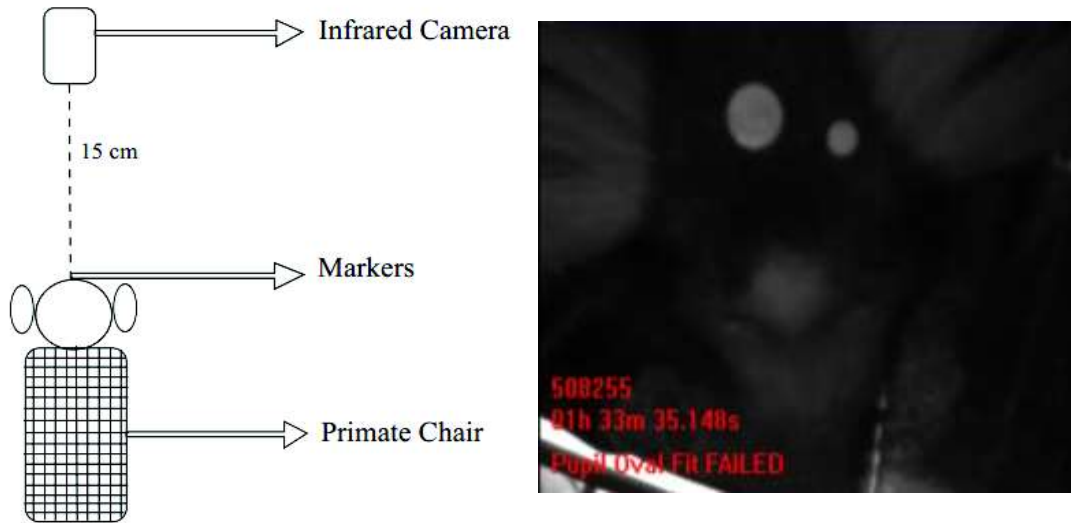
Figure 7: Marker Positioning and a Snap-shot of Monkey Head Movement.

Left panel shows the side view of the monkey in the chair and the camera. The right panel shows the image captured by the camera and the markers on top of the monkey head.

Each trial of the experiment consisted of a set of stimuli presented as individual events with a time gap of around 15 – 60 seconds between each event. Depending on the stimulus the entire trial could last for around 2- 7 minutes and the video recording started earlier and ended later than the stimulus presentation to capture the whole response sequence. The average file size was between 100 and 1000 megabytes in the AVI format.

Each video file was further cut into individual short segments for further processing. Segments that were chosen all contained at least one head motion of interest. These video files were processed using the Windows Movie Maker software and each file was roughly 300 frames in length or 3 seconds in duration. The file was converted the mp4 format since it is compatible with both windows and Mac operating systems and has good compression properties with smaller video files.

In addition to the high-speed camera to capture the head motion, another infrared camera which was used to record the body movement of a monkey in chair at a lower frame rate (30 frame/sec). This was needed to monitor the general behavior of the monkey, to check whether or not the monkey was in a relaxed state. The additional camera was also used to provide information of body orientation for validating the results of the experiments.

There is an Arduino-Matlab setup that gives the control of the setup to the experimenter. Arduino is a microcontroller and here it was used to drive LED lights. A Matlab GUI is presented to the experimenter where there are options for different stimulus and background conditions. Since the video being recorded contains only the information of images captured by the camera and no audio content, there was a requirement of external setup to indicate the onset of audio stimulus. This was done by placing the LED in the field of view of the camera. The LED was fired via the Arduino and the video recording clearly shows the LED glowing.

3.3 Selection of Markers

Two circular markers were used to indicate the head positions that were captured by the IR-camera. The markers were placed on top of the head of the monkey. The two markers had different sizes (the large marker is of diameter 5 mm and the small marker is of diameter 2 mm).

Ideally, three or more markers would be better in obtaining spatial positions. The difference between the two- marker and the three-marker system is that the three-marker system would result in three angles as opposed to the one angle returned by the two marker system. This adds no additional information to the system but would be more

robust. In experiments, it was observed that for horizontal rotation, the two-marker system produced similar results as the three-marker system.

Many options were considered for selecting appropriate markers. Figure 8 shows the marker materials that were tried and their presences in the image. The first attempt was artificial eye parts in order to benefit from the eye-tracker capacity of the Arrington system (Fig. 8A). However the Arrington software was not able to detect the artificial eyes at all times because the plastic casing of artificial eyes generated unwanted glares in images when the head was rotated to certain angles but not all. The second attempt was to use reflective tape (Fig. 8B). However, the difficulties occurred because it is time-consuming to make the reflective tape material into desired circular shapes and to attach them to the monkey head.

The third attempt was flat beads used in crafting (Fig. 8C). They were considered because they come in a variety of size and have perfect circle shape. However, similar to the problems with the artificial eyes, they yielded unwanted glares under the Infra-red camera, and these glares made it difficult to trace the marker position.

The fourth attempt was the medical tape that was adhesive on one side (Fig. 8D). The tape was easy to be attached to the monkey head and the imaging result was good. However it was not easy to cut out circles from the tape.
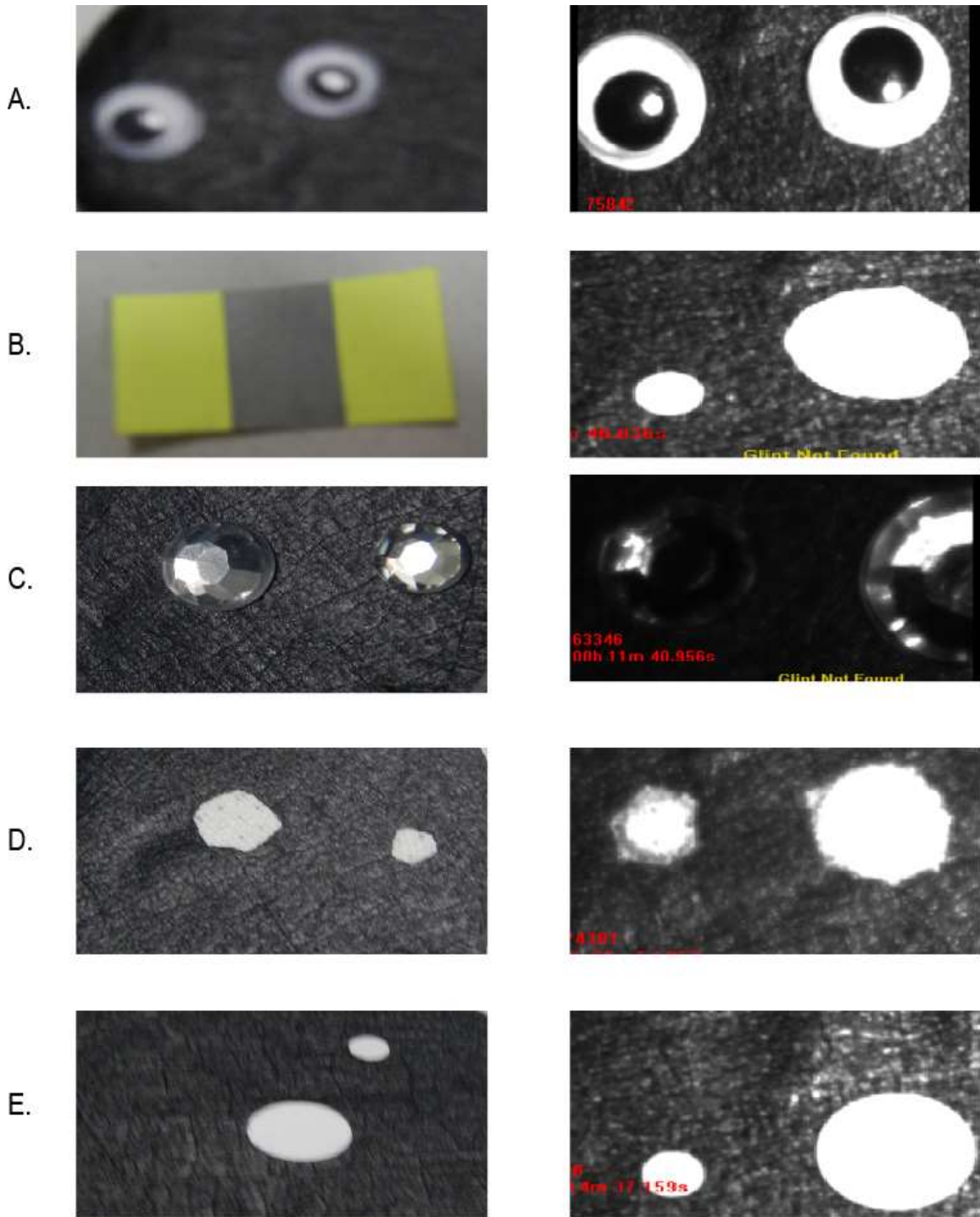
Figure 8: The different Types of Markers Tested in the Experiment.

A.Plastic eye, B. Reflective tape, C. Reflective silver bead, D. Medical tape, E. White cardboard, Left panels show the markers in normal light and the right panels show the same group of markers in infrared light.

Based on the above experience, our final choice was surprisingly simple. Small round disks in different size were cut out from a thin cardboard material using punchers for crafting. Medical ointment was used to stick the disks to the monkey head. As it can be seen from Fig 8E, the disks have perfect round shape and the imaging results are as good as those from reflective tapes.

3.4 Stimulus and Conditions

The head movements of monkeys were examined under various testing conditions with auditory or visual stimuli or both. The auditory stimuli included monkey calls like phee and twitter, human whistling sounds, and nature sounds downloaded from the Internet. The monkey calls were generated using the virtual monkey call program (DiMattina & Wang, 2006). The user interface for stimulus presentation provides complete control about the stimulus properties that were varied, including signal attenuation, background noise attenuation, and stimulus duration. Stimulus parameters can be changed from trial to trial.

Figure 9 shows the graphical user interface (GUI) that was built in Matlab for the purpose of generating auditory stimulus. There was an option for selecting the monkey's name for the purpose of keeping a record. The experimenter had the choice of selecting one out of three stimulus types (twitter, phee or broadband noise). Also experimenter could select whether sound was to be presented on let speaker, right speaker or both speakers. The values for stimulus and background noise attenuation could be entered. The experimenter hits "start" for presenting a stimulus with the properties given above. After completing a trial, the experimenter hits "save" and then "exit".

28
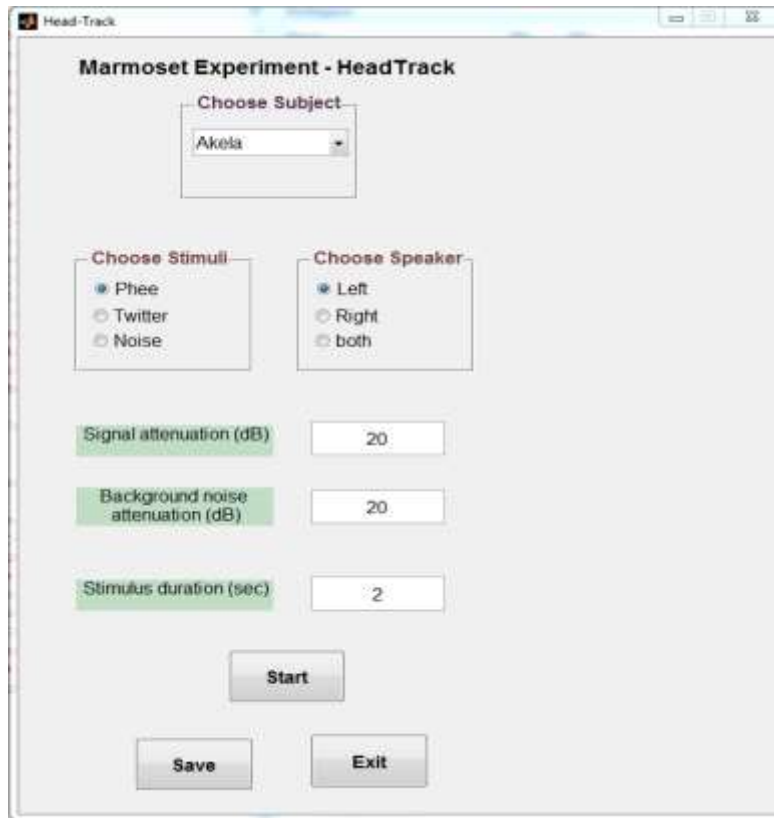
Figure 9: Graphical User Interface (GUI) Developed in Matlab for the Purpose of Stimulus Presentation.

The visual stimuli included monkey images and video files downloaded from Internet.

Monkey behaviors were closely monitored during the stimulus presentation and stimuli were frequently shuffled to avoid adaption and fatigue. For data analysis, the following set of conditions is considered:

- Idle state: Here there is no stimulus presented. The monkey is sitting in the chair with the door of the acoustic chamber closed.

- Sound only stimulus: Here there is sound presented on one or both of the speakers. There is no visual cue. The sound could be twitter and phee calls, bird

chirps, forest sounds, and human whistling sounds. For simplicity they are grouped under one category.

- Visual- Sound stimulus: Here visual cues are presented along with audio cues. The video/images are projected onto the screen in front of the monkey. The images could be those of the colony, marmoset videos (not of the colony), slow strobe light, and person in PPE.

- Alert stimulus: Interesting and consistent head turns are observed for opening and closing of the acoustic chamber door and this stimulus is included in this research. This is an ecologically viable stimulus since in the natural habitat, there would be many such alarming conditions.

For each of these categories, the house lights could be on or off and therefore the total number of categories for analysis is 8. To distinguish between different types of audio stimuli, we make use of an encoding scheme and the encoded information is saved at the end of the experiment as "stimulus order".

Understanding marmoset calls is useful for this research. Marmosets generally make high pitched calls. In their natural environment, vocalization is very important since it may be difficult to visualize a marmoset in thick vegetation. These calls may aid in communication within group members as well as between groups.

The parameters that characterize a call are the frequency with which the call is made, pitch of the call, loudness of call and context of the call.

The vocalizations can be used to communicate important information ranging from group location, warning about a predator, contestants involved in aggressive encounters, mate

attraction, parents and offspring interactions, etc. (20, *Behavior of the common marmoset, nd*)

Following are the types of calls

- Trill- This is a pleasant sounding call, characterized by cyclically changing frequency. This call can be of any length. The marmoset can make this call without opening its mouth. It is a within group call. It could be a contact call, to identify where the other group members are located (14, *Marmoset Vocalizations, nd*).

- Chirp- This is a pleasant sounding call, consisting of rapid, periodic notes from high-frequency to low frequency. The mouth is slightly open. It is a within group call. It is made in a friendly context, say when favorite food is near-by.

- Twitter- This is also a pleasant sounding call, rapid, periodic notes from low-frequency to high frequency. The mouth is open when this call is made. Can be a within group or between group call. This is a territorial call.

- Phee- It sounds like a soft whistle; it is an almost constant pitch call. It may be made in succession (ranging from a series of one to several in a burst). It is a within group contact call.

  Other types of calls are Chatter, Ek, Tsik, Seep.

In the current research, twitter and phee calls are extensively used. The comparison for different types of stimulus is given in chapter 5.

CHAPTER 4: DATA ANALYSIS, CALIBRATION AND VALIDATION

The following section shows the ways in which algorithms described in Chapter 2 was implemented and related calibration methods used in the current project.

4.1 Marker Detection

The central component of the data analysis was to detect the head movement using the kinetics of markers extracted from still images. The blob detection code was used to detect markers in images containing head movement. The entire results and conclusions of this thesis were based on this method.

The data were processed in three steps (1) Pre-processing; (2) Marker detection and (3) Rotation angle calculation.

Figure 10 shows the results of each step of this method in detecting two markers in one frame. As detailed below, the markers were distinguished based on constraints in the blob area and blob circularity. Two markers of different size were used. The diameter of the large marker was 5 mm and the diameter of the small marker was 2 mm.

Step1: Pre-processing-

As the first step of preprocessing, the RGB image (Fig. 10A) was converted to a grayscale image (Fig: 10 B). The next step of preprocessing, contrast enhancement, was performed by mapping intensity values from a given input range to a given output range (Fig: 10C).
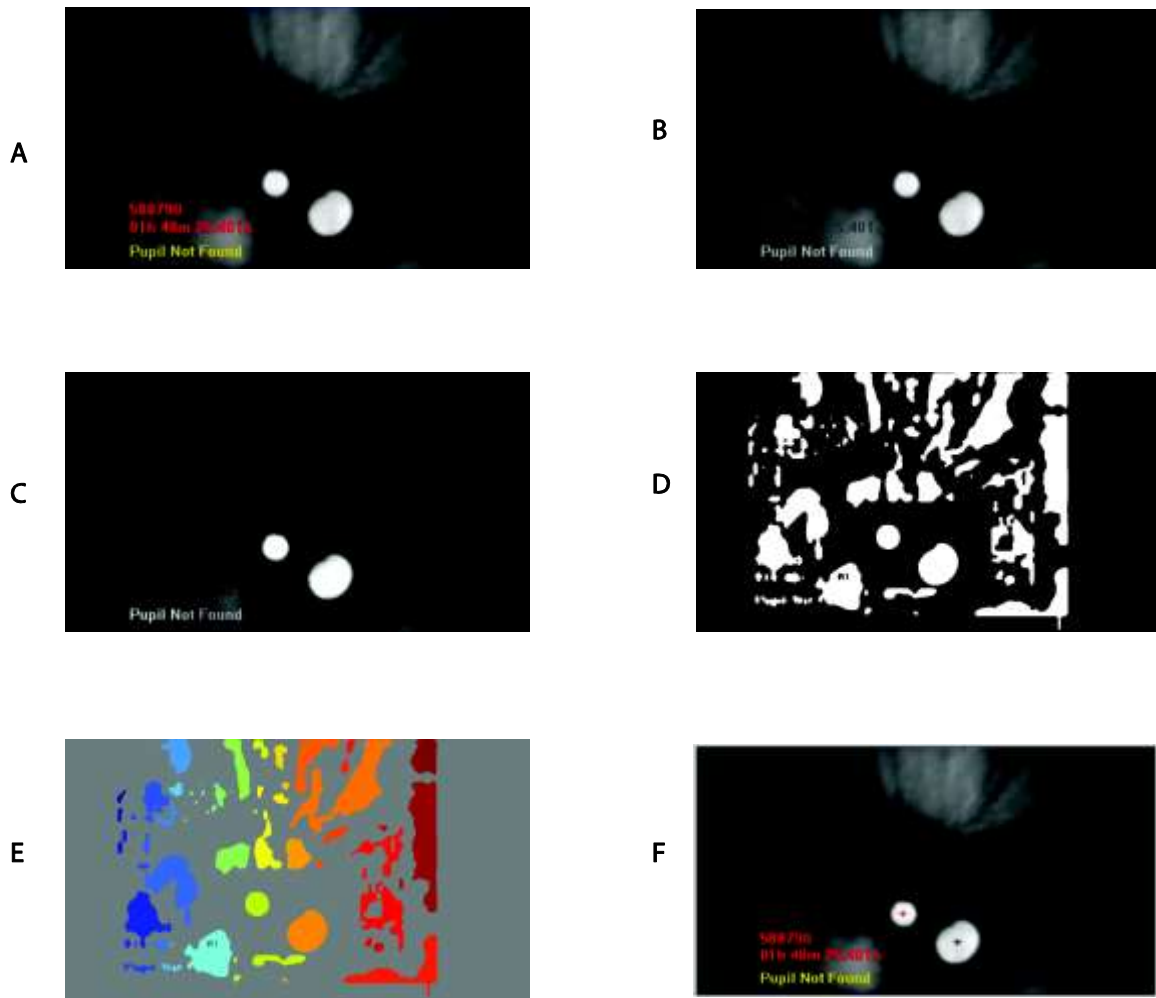
Figure 10: Images Obtained from Different Stages of Processing in the Blob Detection Algorithm.

A. Original image, B. After converting to grayscale, C. After contrast enhancement, D. After thresholding by Otsu method, E. Identifying different regions (blobs) F. Object detected shown as red and black dots.

The grayscale image was converted to a binary image, where the process is called image segmentation. A threshold was required for separating the image into '1' pixels (white/foreground) and '0' pixels (black/background). One way of doing segmentation was to fix a constant numerical value based on trial-and-error methods. This is a global threshold applied to the entire image. Figure 11 shows the result of applying this method.

The Otsu method described in chapter 2 computes the threshold in such a way that the optimum threshold separating the foreground and the background so that the intra-class (classes being foreground and background) variance is minimized (Otsu, 1979). This method can also be extended to multi-threshold situations. In an ideal image histogram there would be a deep valley separating the foreground and background. In real world images, this is not the case.

The Otsu method gives a global threshold k as in equation (1). Although segmentation by this method is optimal, it may not ensure that the markers are detected. Instead of such global threshold value, adaptive thresholding can also be used where threshold is not a fixed value and is local to regions of image. In the adaptive method, a threshold is selected for every pixel. The assumption is that small regions in the image have uniform illumination. This takes into account the variations in lighting conditions in the image. For images obtained in this experiment, the adaptive method is particularly useful since the illuminator settings and the ambient and LED lights for each trial of experiment may vary. Compared to the constant threshold method, the adaptive method avoids to some extent using "trial-and-error" for each video file. Another benefit of the adaptive threshold method was to avoid image saturation from strong signaling light from LED. For the frames that contain LED signaling light, markers could not be detected if a fixed threshold value was used.

The Otsu's method was implemented on blocks of the images to obtain adaptive thresholds. The result was each block would get a threshold value. The block size has to be chosen appropriately, it was chosen to be around 70x70 pixels.

As the final step of preprocessing, smoothing and noise reduction were performed on the binary image using a median filter based on a non-linear filtering technique. The median filter replaces the pixel under consideration with a value given by the median value of a window around a pixel. A two-dimensional median filter was used in the data analysis. For the boundary condition, zero padding was performed as opposed to the repetition method (repeating the boundary values at the ends of the window). The result of smoothing can be seen in Fig. 10D.

Figure 11: Resulting Binary Image at the Image Segmentation Step with Segmentation Performed Using a Fixed Global Threshold for all Pixels in the Image.

Step2: Marker detection

In the marker detection step, the core algorithm served to distinguish markers from other features and the background (10 E). Boundaries of blobs were detected from the resulting binary image after the Otsu thresholding and median filtering. Properties of each blob, such as area, centroid, perimeter boundary coordinates were stored for further analysis. Since markers were circular in shape, the ratio of circumference squared to the area, which was constant in principle $(2*\pi*R)^2/ (\pi*R*R) = 4* \pi$, was used to distinguish the markers from other artifacts in the image.

The first frame was used to determine the properties of the markers based on which the subsequent frames of the markers were analyzed. From the first frame, the area of each marker was noted and the area values for subsequent frames were determined. Four bounds for area size were implemented, a pair of upper and lower bounds for each marker. In the subsequent frames, for a blob to qualify as a marker, it had to pass the criterion for circularity and also the criteria for blob area based on these bounds (i.e. greater than the upper bound and less than the lower bound).

The distance between the markers in the first frame (called base side-length) was noted and used as a conditional parameter for rejection of a pair of markers in the next frames. If the distance between the markers for subsequent frames was found to be greater than twice the base side-length or less than half the base side-length, the detections were discarded as invalid detections.

The bounds for marker area were checked for each frame. If one or more markers were missing, the algorithm checked for blobs in the same region as that of the detected marker in the previous frame. After analyzing the entire length of the video, the criterion for base-side-length was examined for validity of detection.

Step3: Rotation angle calculation

After obtaining the valid marker centroid co-ordinates for all the frames, the head rotation angles in horizontal and vertical planes was calculated using trigonometric analysis. The rotation vector was indicated by the vector connecting the large and small markers. The placement of the markers was not arbitrary. For the purpose of improving the vertical rotation accuracy, the large marker was placed at a point on the monkey's head roughly above the neck which was assumed to be the pivot for the head rotations.

36

Figure 12 shows the calculation schematic. For the horizontal rotation, the angle subtended by the line joining the markers and the y-axis of the image was computed. The Y axis was the zero reference in the coordinate system used in the analysis.
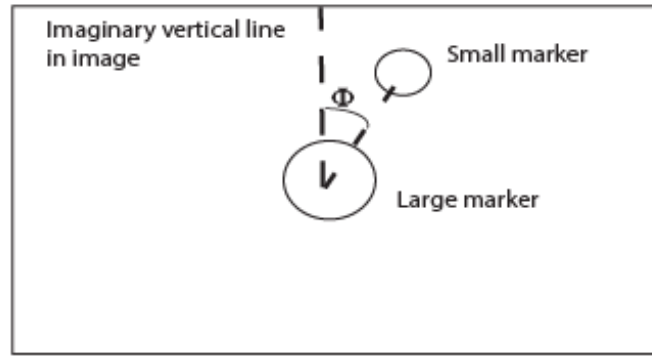
Figure 12: Method for Calculating the Horizontal Rotation Angle.

The angle subtended by the line joining the centroid of the markers and the imaginary vertical line of the image.

The x and y coordinates of the detected markers and a trigonometric identity like inverse cotangent was used in rotation angle estimation. Clockwise rotation was marked between 0 to 180° and anticlockwise rotation was marked between 0 to -180°.

In calculating the vertical rotation angle, the reference frame was first identified. The reference frame is the one where the large marker was directly below the camera and the top of monkey head was most close to be horizontal plane with minimum tilting. The reference frame was the one on which the length variable L (Euclidean distance between the two markers) was at maximum among all frames collected. The length variable L of each frame was then compared with that of the reference frame (ML). The angle in degrees was given by the cosine of the ratio of these lengths. The ambiguity of the positive and negative angles may be solved by comparing the location of the markers

with the location of markers in the maximum frame. This method, however, was not perfect, the vertical rotation component needs to be further investigated in future research. Missing markers in frames also adversely impacted the detection of the vertical rotation component.
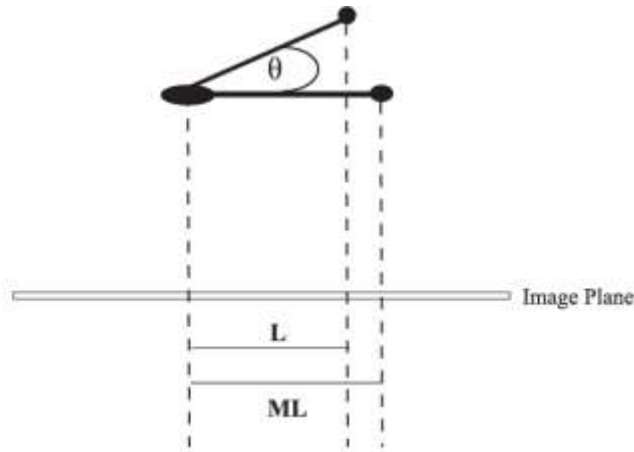
Figure 13: Method for Calculating the Vertical Rotation Angle.

The vertical angle is given by the ratio of maximum distance between the markers to the distance between the markers in the current frame.

The maximum length (i.e. distance between markers) ML and L which is length in current frame are considered as in Figure 13, i.e. Vertical rotation, $\theta \propto \frac{L}{ML}$

Again a trigonometric identity like cosine is used. To increase the accuracy of this method, the maximum length can be computed for the entire trial of experiments (i.e. for the entire raw video file) and stored and for each of the individual trimmed video segments this maximum length value can be used.

Figure 14 shows the reason why this method is accurate for determining the horizontal rotations. It can be seen that the relative horizontal angle of the two markers with respect

38

to the horizontal of the image is unchanged when subject to translational motion or vertical rotational motion.

A

B
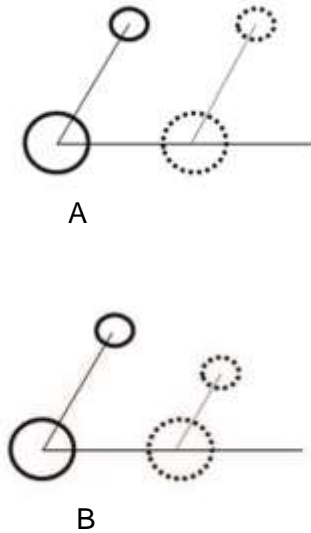
Figure 14: Effect of Translation and Vertical Tilting on Detection of Horizontal Angles.

Panel A shows that translation alone would not change the calculated horizontal rotation and Panel B shows vertical rotation alone would not change the calculated horizontal rotation.

4.2 Optical Flow

This is a motion detection method and was considered in this thesis to provide additional methods for head movement detection. This method was not used to obtain the results and conclusions and the reason is given in chapter 6. The algorithm version used in the current scenario is of affine flow (Young, D. 2010). It is based on the Lucas Kanade method. This is a sparse optical flow method and computationally less expensive than Horn Schunck method as mentioned in the book by Bradsky & Kaehler (2008). The affine flow field model used here has 6 parameters made of following variables: rate of

dilation, rate of rotation, shear along image axis and shear along diagonal axis and optical flow at the origin. Algorithm performs a least square fit to the model by calculating the spatial and temporal gradients for the image sequence taking two at a time.

Some of the parameters for the optical flow object are, type of spatial sampling-rectilinear or logpolar, region of interest, smoothing filter parameter (Gaussian filter is used), if rectilinear sampling is used, a sampling interval can be selected as sampling need not be done for every pixel. For this particular analysis, a sampling interval of 10 is chosen. This should give a good balance between the accuracy and the computation time. Also a Gaussian mask of 10*10 is chosen to perform smoothing. Figure 15 shows the optical flow fields for these parameters in one frame of the monkey video.

As an extension to the Lucas-Kanade method, given in the Bradsky & Kaehler (2008) book is the pyramidal LK method where the degree of motion is taken into consideration. This has additional steps when compared to the LK method. This is done by considering different scales of the images and the algorithm is started with coarse set first and then to fine flow computation.

Figure 15: Optical Flow Fields in One Frame of the Video.

4.3 SURF

This is a feature detection method and was considered in this thesis to provide additional methods for head movement detection. This method was not used to obtain the results and conclusions and the reason is given in chapter 6. The theoretical basis of this algorithm was given in chapter 2. In the present analysis, first binary image segmentation is performed. Prior to this median filtering is executed on the grayscale image. Then a certain number of strong features based on SURF are detected and are extracted into vectors. The subsequent frames are subjected to the same steps and the features are compared with those of the first frame and features are matched. Based on the matched features, the rotation angles are determined using affine transformation equations. Intermediate steps of feature detection and feature matching are shown in figure 16.
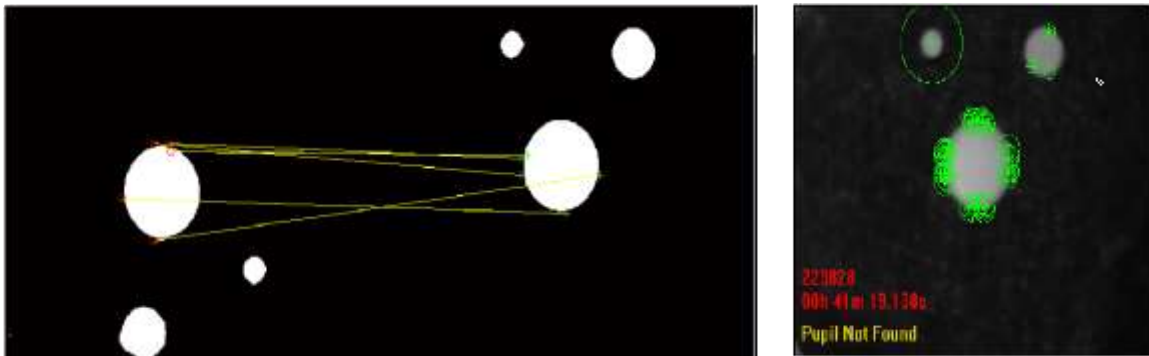
Figure 16: SURF Feature Detection and Matching in one Frame of the Video.

Left image shows the features being matched between the first frame and current frame of video. Figure on right shows the best features in one frame.

4.4 Kalman Filtering

This is an object tracking method and in this research it was used in conjunction with the blob analysis algorithm. The entire results and conclusions of this thesis used this

41

method. As discussed in chapter 2 there are three categories of system models and the best possible type of system that fits our needs here is possibly the random motion model. This is a model to be used when the object has the ability to make movements on its own. This kind of a system works in such a way that the variances in the state estimates are increased. As an alternative, the constant acceleration model was chosen since the object (here markers) are in the field of view of the camera for a considerable range of movements the monkey can make. Also the linearity assumption was put to test in the present scenario. In the video segments of the marmoset head turn, there are sections where the motion was absent and there are sections where motion was very fast.

Also it must be noted that since the accuracy of the measurements with the blob analysis algorithm was high, especially with the additional checks on detection validity, the marker locations where detection occurred are not updated and the Kalman filter was used to estimate the position of the occasional missing marker. The algorithm does not make use of the correction phase of the Kalman filter, only the prediction phase was used to estimate this missing marker. The following were the reasons for the marker to go missing:

1. The marker circularity criterion fails because of motion.

2. The marker was at the edges of the image of at the image artifacts (like the text in the bottom left corner) and because of this, the marker circularity criterion or the marker area criterion or both fails and the marker was not detected.

3. The head movement of the monkey was too fast, even for the high-speed camera (with frame rate of 90 fps), and because of this there was motion blur in the image and the markers in the image were not circular and got rejected.

4. Because of some of the image processing techniques used (like median filtering, contrast improvement), sometimes the marker may not be detected.

5. Sometimes the length criterion (distance between the markers) fails because an artifact is considered as marker.

The following discussion is taken from the book by Bradsky & Kaehler (2008) to explain briefly the mathematics of the Kalman Filter. Consider a linear function F which is applied to the objects' state. Consider the observation model Z where the parameters of this model consist of some of the variables of the system state. The remaining state variables depend indirectly on the observation model parameters.

Let x_k be the n-dimensional state vector at instant k and F be the n*n matrix called the transfer matrix. w_k is the process noise which has a Gaussian distribution. The relation with z_k which is m- dimensional (generally m>n) is also given where H_k is the m*n dimensional matrix and v_k is the measurement error.

$$x_k = Fx_{k-1} + w_k \tag{1}$$

$$z_k = Hx_k + v_k$$

First the update step is performed by computing the priori state vector estimate

$$\bar{x}_k = F\, x_{k-1} + w_k$$

The equation for the a priori error covariance estimate is given by

$$\bar{P}_k = F\, P_{k-1} F^T + w_k + Q_k$$

where Q_k is the process noise covariance.

43

The Kalman gain is given as

$$K_k = \bar{P}_k H_k^T (H_k \bar{P}_k H_k^T + R_k)^{\wedge}-1$$

where R_k is the measurement noise covariance.

Now the correction step can be written as

$$x_k = \bar{x}_k + K_k(\bar{z}_k - H_k\bar{x}_k)$$

In this study, Kalman filter is applied to the result of Blob detection algorithm after the detection validation (depending on base side length) is performed. At this step, there is a matrix of marker locations with gaps in the frames where there is no detection or where the detection has been discarded as invalid. Now for each marker, the Kalman filter was initialized and the prediction and correction computations for the entire length of the video frames are performed.

For the constant acceleration model, the parameters in equation (1) are,

$$F = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Here x_k is the location of the marker. Note that the individual dimensions of the position vector are considered separately i.e. the analysis is done disjointedly for the x and y coordinates of the markers.

As mentioned earlier, the Kalman prediction results are used to "fill-in" the spots where there was no detection. After this step was performed for both markers, the horizontal rotation was computed as in the blob detection algorithm. An alternative to this method is to compute the horizontal rotation first and then perform Kalman filtering on this rotation angle.

4.5 Detailed Algorithm

Below is the summary of the algorithm used to identify markers and calculate the rotation angles. Figure 17 shows the corresponding flowchart.

1. Open the video file using the VideoReader object of the Computer Vision toolbox in Matlab. This class gives us certain important properties of the file such as number of frames of video, width and height of individual frames.

2. Read the image at frame 1. Convert the image from color to gray scale. Using Otsu's method, perform binary image segmentation on image blocks of size 70*70. Perform median filtering on the resulting binary image using 'medfilt2'.

3. Using the 'bwboundaries' function determines the different boundaries in the frame. Very small regions can be eliminated from further consideration. For each of these regions, determine the area and centroid using 'regionprops' function.

4. Loop through each of these regions; determine which region meets circularity criterion. There are two such regions corresponding to the two markers. Set marker area bounds from the area values of these regions.

5. Calculate the distance between the centroid of the markers. This gives the "base side length", the reference side length for validating detection.

6. Loop through remaining frames. Repeat steps 2 to 4. In step 4, check for blob area and circularity criterion. Extract the marker centroids in the X-Y coordinates display them on the original image. Mark the vector (line joining the smaller and larger markers) on the "marker image" figure.

7. Eliminate the wrongly identified marker and miss marker frames using the length of the vector as a criterion. Perform Kalman filtering to determine missing marker positions. Calculate maximum side-length (distance between markers)

8. Compute the horizontal rotation angle by finding the angle subtended by the vector with the vertical. The algorithm uses the 'acot' function and the angles can be scaled and shifted depending on the quadrants.

9. Compute vertical rotation angle from current frame side length and maximum side length. Use 'acos' function. Change the magnitude depending on position of markers with respect to previous frame.

10. Calculate the velocity(in degrees per second) and acceleration (in degrees per second squared) using 'diff' function in Matlab.

```
┌─────────────────────────────┐
│ 1. Open video file using the│
│    Video File Reader        │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 2. Convert first frame to grayscale │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 3.Perform contrast enhancement │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 4. Otsu threshold on sub-blocks │
│    of frame                 │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 5.Median filter the resulting image │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 6.Obtain regions in resulting image; │
│ save area, centroid and perimeters │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ 7.Compare circularity of all regions;get │
│ porperties of two objects that meet the │
│ criterion; set area bounds for markers │
└─────────────────────────────┘
              ↓
         ◇ Loop through
           remaining frames ◇
              ↓
┌─────────────────────────────┐
│ Perform steps 2, 3, 4, 5, 6 │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Check for area and circularity │
│ constraints; save centroids of markers │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ If no detection look in nearest neighbourhood │
│ of frame containing detection │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Determine maximum length    │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Check length constraints for markers │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Perform Kalman filtering on the │
│ marker centroid locations   │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Compute horizontal and vertical rotations │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│ Compute corresponding velocity and │
│ acceleration by differentiation │
└─────────────────────────────┘
```
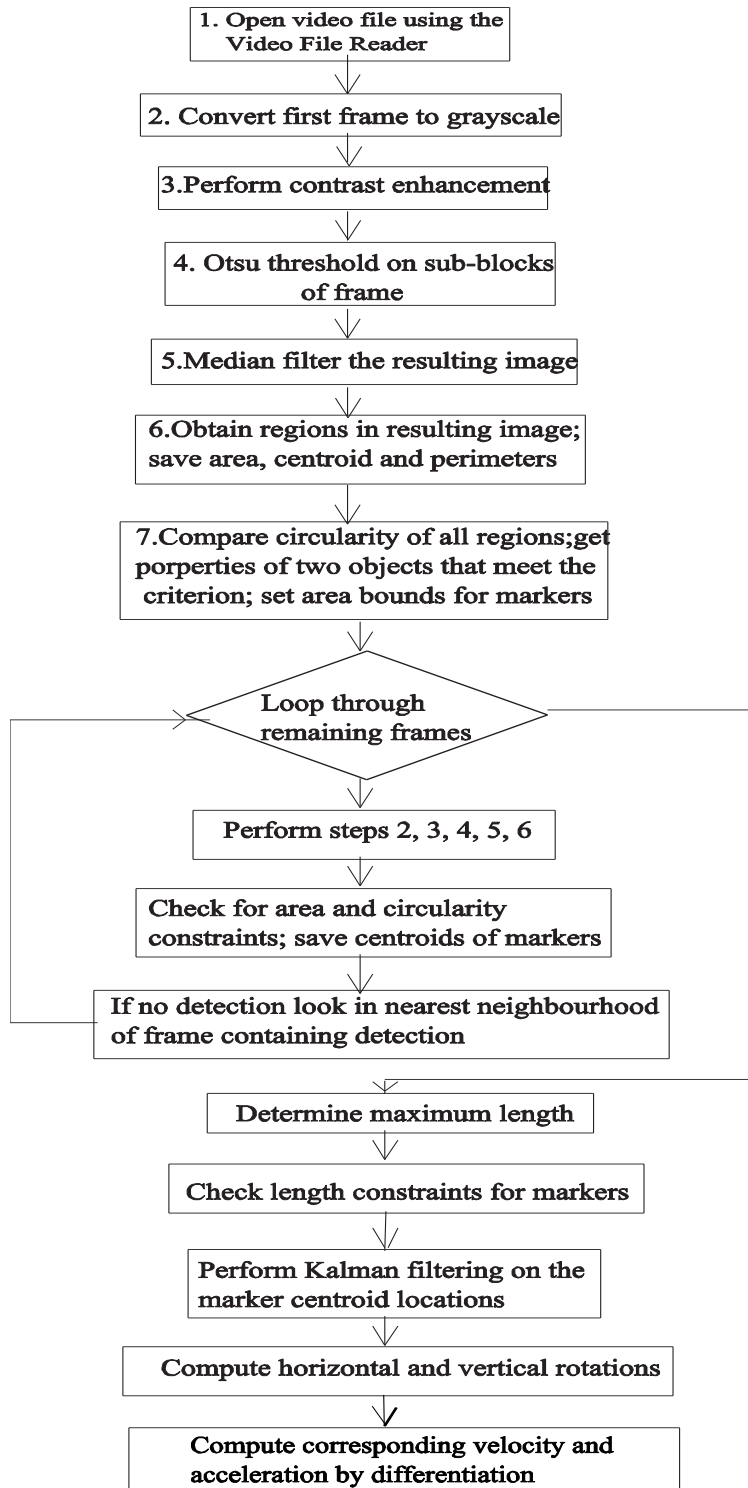
Figure 17: Flow Chart for the Detailed Steps in Implementing the Blob Detection –
Kalman Filter Algorithm.

4.6 Calibration

Ground truth data was used to compare the accuracy of different methods discussed above. Ground truth data composed of hand-marked marker position data using the Matlab function 'ginput'. They were performed on the video segments randomly chosen. For a chosen video, the markers were marked on the image for every other 5 frames. The computer program stored the coordinate information of the markers obtained.

For the purpose of validation of results, a 12-V electrical motor system that yields constant angular velocity was used to test the validity of the data analyses method. Figure 18 shows the calibration setup. A lightweight disc made out of foam was attached to the motor shaft. The markers were attached onto the front and side faces of the foam disk. The motor body was placed on the chair at about the same height as the monkey head. The angular velocity of the motor computed from the algorithm was compared with the actual velocity of the motor. It was confirmed that the error of the computed velocity was within +/- 10 degrees/sec as compared to the actual velocity ranging from 200 degrees

Figure 18: DC Motor Used in Validation and Calibration.

Left image is the setup used for measuring horizontal rotations and right is the setup used for measuring vertical rotations.

per second to 400 degrees per second. The calibration system has the advantage of separating the horizontal rotations and vertical rotations with no translational motion.

CHAPTER 5: RESULTS

This chapter reports the results of marker detection and head movement obtained in the experiments. The first part reports the results comparing different methods using the methods discussed in Chapters 3 and 4 and the second part reports the results of monkey head movements tested in different conditions.

5.1 Comparisons of Results Using Different Marker Detection Methods

Fig. 19A and 19B show the time evolution plot for the XY position of the markers i.e. it is a frame by frame representation of the marker positions all consolidated into one plot. Fig 19A is for motor video data and Fig19B is for monkey video data. The corresponding angle plots for the data are given in Fig. 19C for the motor data and Fig. 19D for monkey head rotation data. Note that as expected the horizontal angle for motor data is a straight line, the motor rotates from 0 to 360° with a constant step size. The breaks in the curve indicate jumps from 360° back to 0° for the next turn. The slope of this straight line would give the angular velocity. Since the expected result was achieved for the motor data, the same algorithm was applied for the monkey head rotation data with confidence and the corresponding plots are seen in Fig 19 B and Fig 19D.
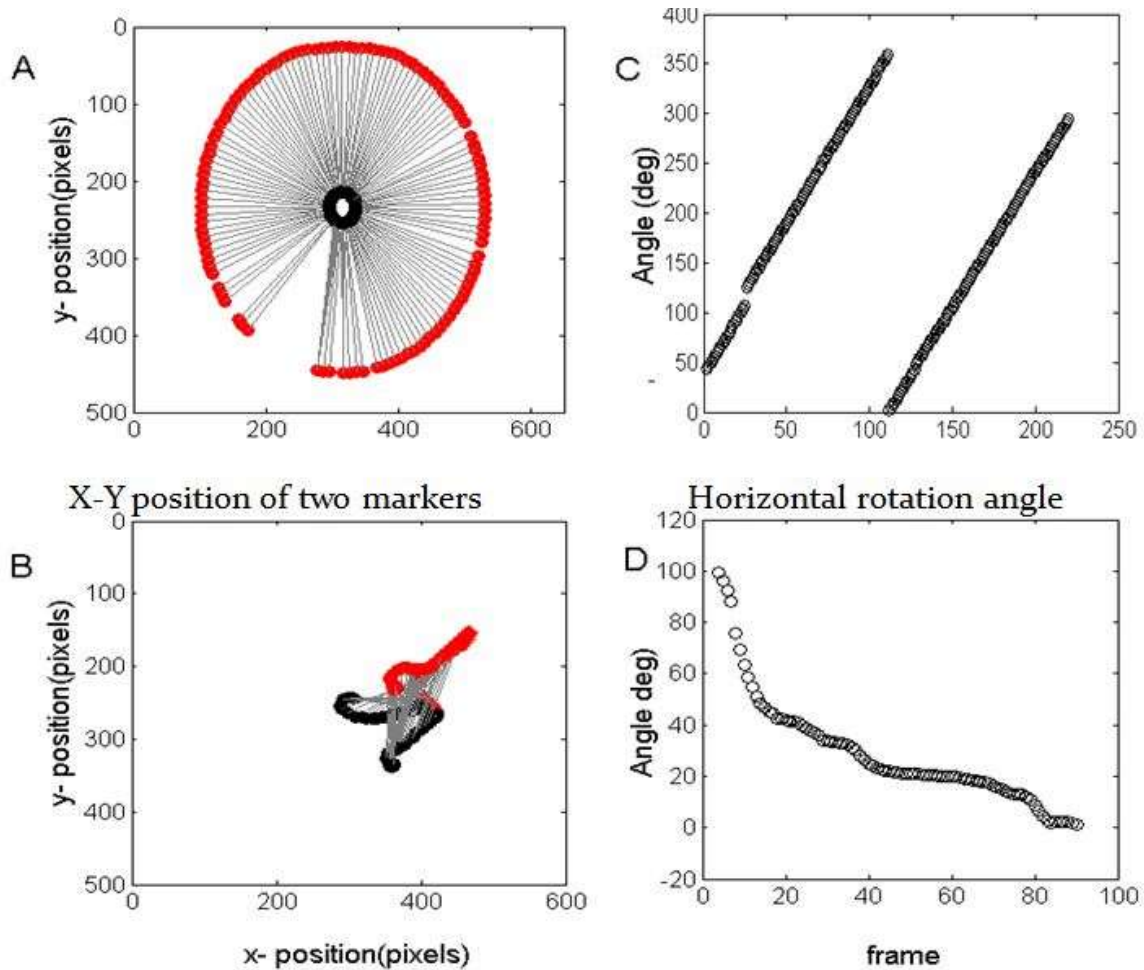
Figure 19: Time Evolution Plot for x-y Positions and Corresponding Horizontal Angle Plots of the Two Markers for Both DC Motor Video and Monkey Head Rotation Video Data.

A. X-Y position of two markers captured when the DC motor made a 360° turn, B. X-Y position of two markers captured during monkey head rotation, For both A and B, red circles show the time evolution of the positions of the small marker, whiles the black circles show the time evolution of the positions of the large marker C. Horizontal rotation angle plot for DC motor data corresponding to panel A, D. horizontal rotation angle plot for monkey head rotation data corresponding to panel B.

Figure 20 compares the actual and measured velocity of the rotation of the DC motor.

Velocity was obtained by differentiation of the horizontal angles of head rotation. Bursts

in measured velocity indicate spurious data points, at which frames markers ran out of the

51

sight the camera. It can be seen that the calculated velocity of the horizontal rotation of the DC motor matched well with the actual value. For vertical rotations, as discussed in Chapter 4, it was difficult to obtain vertical angles from monkey head movement. However, for DC-motor rotation, the analysis worked well, indicating that in principle the decoding algorithm could detect vertical movement. It was also able to accurately give positive and negative angles (Figure 21).
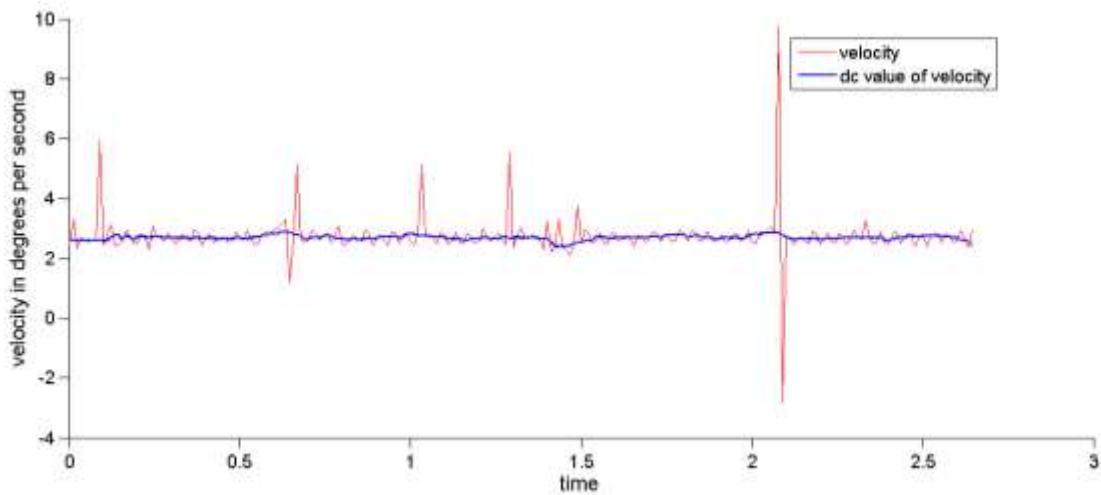
Figure 20: Horizontal Velocity Plot When the DC Motor Made Two Full Turns (720 Degrees).
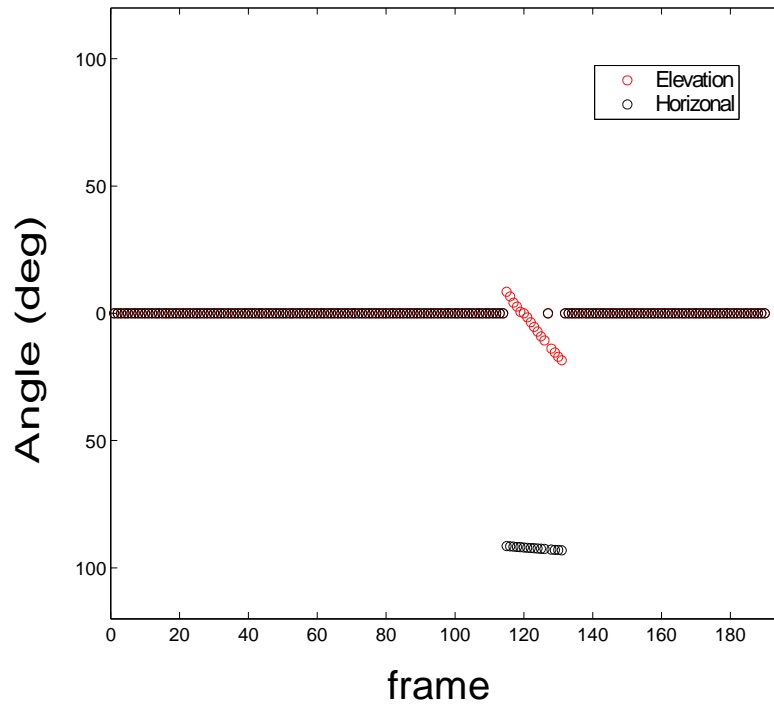
Figure 21: Vertical Rotation Angle When the DC Motor Made One Turn (360 Degrees).

Figure 22 compares the results from four methods in detecting the horizontal movement of a DC motor. The red symbols show the actual values of horizontal angle based on the ground truth data and the blue symbols show the results of different algorithms. The DC motor rotated one full turn in 360 degree. It can be seen that the blob detection algorithm performs well except missing markers in two frames when the markers ran out of the field of the view of the camera (Fig. 22A). As shown in (Fig. 22B), the optical flow algorithm does not perform as well as the blob detection. One possible explanation is that the markers occupied small regions in the image and there were not enough moving points for the optical flow code to work with. This may explain the accumulative errors of underestimation over time in the results.
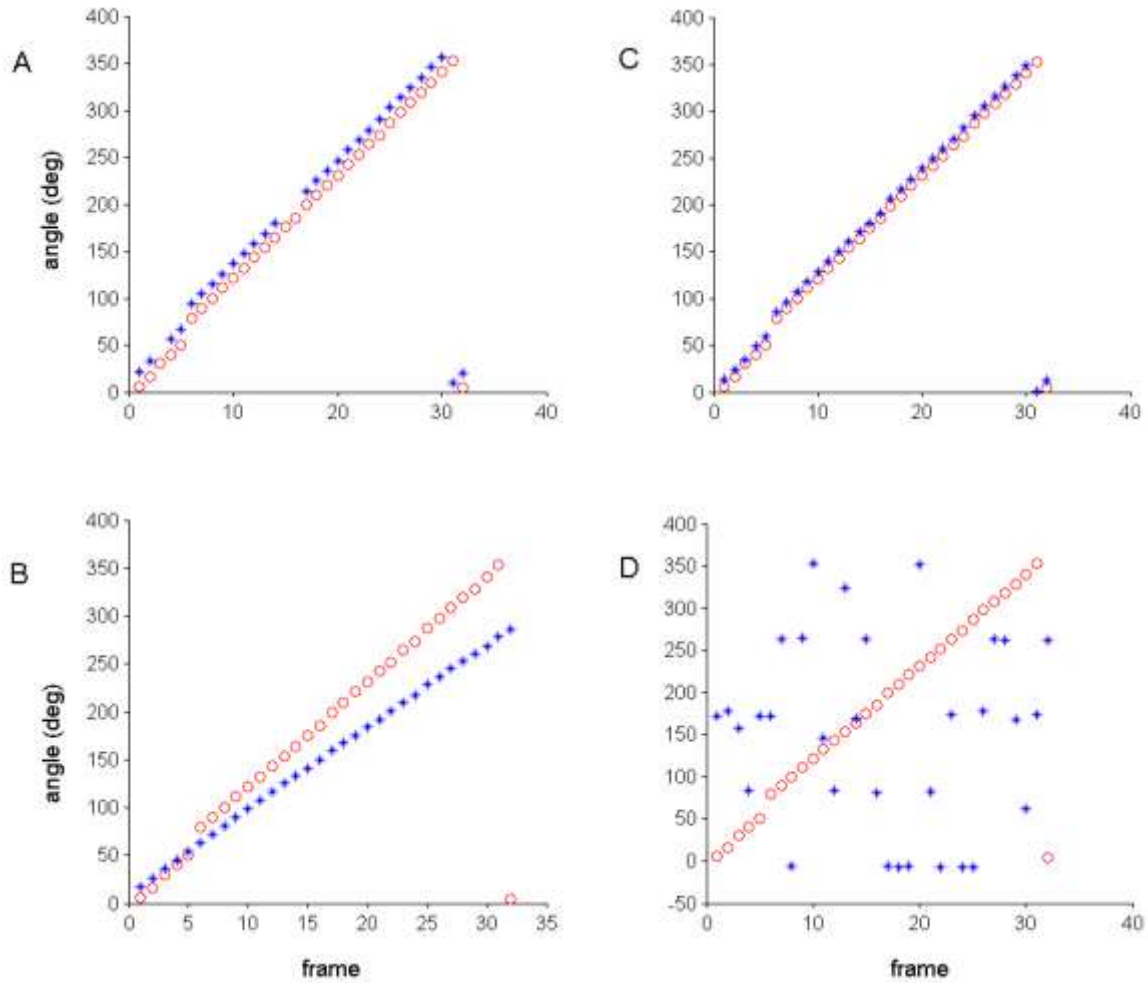
Figure 22: Comparison of the Horizontal Rotation Angles Obtained Using Different Marker Detection Methods When the DC Motor Made a 360-Degree Turn. Red Dots Show the Ground Truth and Blue Dots Show the Results of Different Methods.

A. Blob detection, B. Optical Flow, C. Blob detection with Kalman filtering, D. SURF

As shown in (Fig. 22D), the SURF algorithm performed quite poorly. This can also be

seen in Fig 16 in Chapter 3, where there were a lot of mis-matches in the features

probably because the image did not contain corners. In comparison, blob detection with

the Kalman filter yielded the best performance for this example movement (Fig. 22C). It

54

was mostly true when using constant acceleration model. The same model was not as effective for the monkey video data.

Table 2 gives the Root Mean Square (RMS) error of the different algorithms evaluated in the Chapter 4. The reference used was same ground truth data estimated from the one rotation of the DC motor. Note that for the blob detection case, the error is Nan because missing data points causes RMS calculations ran out of chart.

Table 2: RMS Error Values for Different Methods

Method	Blob analysis with Kalman	Blob Analysis without tracking	Optical Flow	SURF
RMSE angle(in degrees)	1.59	Nan	66.36	123.5
RMSE velocity (degrees per second)	22.37	Nan	56.95	2731.2

Another criterion for comparison is the timing profiles for each algorithm. Again, one turn of the motor data was considered and results are summarized in Table 3. It can be seen that the time for optical flow analysis is very less. The blob detection algorithm takes between two to three times the run-time for optical flow. One reason for this is that some image display functions with high time overhead were used. The SURF algorithm takes a very long time, about five times that of optical flow. Note that the blob detection algorithm with Kalman filter does not have considerable performance overhead for the additional Kalman filter computations. Therefore, for processing of the actual monkey head rotation data, we used the marker detection algorithm coupled with tracking by Kalman filter.

Table 3 : Time Required to Complete the Analysis for the DC Motor Rotation Using Different Methods. The DC Motor Made a 360 Degree Turn.

Method	Total time (sec)
Optical flow	25.43
SURF	137.3
Blob detection	65.56
Blob detection with Kalman filter	65.79

5.2 Results of Monkey Head Movements.

As mentioned in chapter 3, there are four conditions that the data can be categorized into-idle state, sound stimulus, visual-sound stimulus and alert stimulus. For processing and analyzing the video files containing the monkey head rotation data, the parts containing the motion were segmented. The segmentation was done so that the beginning and end of the segment contains idle state (no motion) and such a segment was considered a good segment. There were segments where the initial frames did not contain the idle state (because the markers were not visible in the frame). These segments were not used in the analysis. Also each segment may have multiple head movements; the individual sweeping motion was isolated and used for analysis. Table 4 gives the number of video segments obtained for each condition and the total number of individual motion.

Table 4: The Number of Raw Video Files and Total Number of Segments in These Files for Different Stimulus Conditions.

Stimulus	Trials-light on-condition	Segments-light on condition	Trials-light off-condition	Segments-light off-condition
Idle	5	7	5	10
Sound-only	25	87	13	83
Visual - sound	4	9	8	21
Alert	5	52	4	17

Out of these the numbers of individual motions that were extracted which satisfy the criterion for a good segment are given in Table 5.

Table 5: The Number of Individual Motion Segments Used During Analysis.

Stimulus	Individual motion analyzed(light on)	Individual motion analyzed(light off)
Idle	11	15
Sound-only	22	17
Visual - sound	23	24
Alert	27	15

To obtain Figures 23 through 26, which are the example plots for different conditions, the resulting horizontal angles were stored in a matrix for all the video segments and date were grouped according to the stimulus type. The number of individual motions inside each segment was noted down since a segment may contain more than one motion.

Then the maximum velocity for each segment was computed. The data was arranged in increasing order of maximum velocity. To make comparisons easier, the angles were zero-shifted i.e. all angles are normalized so that they always begin at zero. Also all turns were made absolute, i.e. start at zero and increase to 180 to aid in comparison. Two variables indicating starting phase information in degrees and a symbol indicating whether or not phase change during absolute phase conversion was performed were stored in a "result" data structure.

For each of Figures 23 throiugh26, the panels A through C show the results for house light on condition. Panels D through F show the results for house light off conditions. Panels A and D contain the angle plots, panels B and E contain the velocity plots, panels C and F contain the acceleration plots. In panels A and D, the graphs have a gray dot on

them indicating the global maximum of the curve and in panels B and E, the gray dot represents the local maxima in the curves. There was a threshold for selecting the maxima which was set at 40 degrees per second for all conditions.

For each of the stimulus conditions, four examples are presented in this document which represents the entire dynamic range of the data. Firstly in the Figure 23, for the idle state observe that for both the house light on and house light off conditions, the peak velocities are in the range of 200-600 degrees per second. In these angle plots, a "flat" line segment (i.e. parallel to x axis) indicates a stationary state and the line segment with a slope indicates motion. Notice the differences in the fourth plot in panel A and the fourth plot in panel D. The one in panel A has initial flat segment a "motion" segment and a final flat segment whereas the one in panel D has a flat segment, followed by a motion segment, followed by another flat segment, and another motion segment and finally a flat segment. There is a discontinuity in the head motion, which give the appearance of a multi-phase motion and this behavior is worthy of future investigations. Because of this kind of motion, the velocity plot would have multiple local maxima, as can be seen in panel E, example 4. Another important feature that one can observe from the graphs is the duration of head motion. For Figure 23, note that in Panel A, first and second example, have a longer duration of head turn, lasting for almost 2 seconds and in examples 3 and 4 the duration is close to one second. Thus in general, slow head turns last for a longer time. In some cases like Fig. 23, Panel A example 2, it is difficult to determine where the motion is occurring and also determine the exact velocity since there are many local maxima.

In figure 24, which is the sound stimulus example, from panels A, B and C, for the house light on condition, the dynamic range of head turn velocities is 100- 600 degrees per second whereas for the house light off condition from panels D, E and F, the range is 100-400 degrees per second. In Figure 25, which is the visual and sound stimulus example, from panels A, B and C, for the house light on condition, the dynamic range of head turn velocities is 100- 800 degrees per second whereas for the house light off condition from panels D, E and F, the range is 100-600 degrees per second.

The more interesting and different results are seen in Figure 26, which is the alert stimulus example. For the house light on condition, the dynamic range of head turn velocities is 100- 2500 degrees per second and such large velocity is not observed for any other stimulus. For the alert stimulus, the head turn was always in the direction of the door and there is always a reverting motion associated with the primary motion. For simplicity this is not represented in the current thesis.

 For the house light off condition from panels D, E and F, the dynamic velocity range is 200-800 degrees per second. The examples for the alert stimulus also show both single phase and multiple phase motions, however, from panel A example 3 and 4 which are fast head turns, there is effectively only single phase motion. For the same examples, note the duration of head turn is nearly equal to 0.5 seconds whereas for the other examples, the duration is one second. In panel D, the example 3 has duration of 0.5 second whereas the other examples have duration close to one second.

Zangemeister *et al* (1981) studied the head rotations for gaze and eye-coordination. It can be seen that the results of their research has a similar sigmoidal pattern as in the figure 23. Similar results are seen in the Guitton and Volle (1987) research.

While conducting the experiments, it was observed that for sound and visual-sound stimulus, the marmoset exhibited a "searching" behavior. They move their heads around the room looking for the object producing that sound. This can be seen also form the graphs of Fig. 24 and Fig. 25, where the velocities are small. For the alert stimulus Fig. 26, involving opening and closing of the acoustic chamber door, and the head turns were surprisingly fast.

Since for each of the stimulus conditions, there are both slow and fast head turns, conclusions cannot be drawn from this information alone. A logical step was to perform histogram/summary analysis to observe the number of head turns. Figure 27 and 28 shows the histogram plots for the above results in terms of bins of maximum turn angle and maximum velocity. A similar analysis was performed by Zangemeister *et al* (1981) where they characterized the head rotations based on peak velocity, peak acceleration, time of these peaks, and duration of the movement.
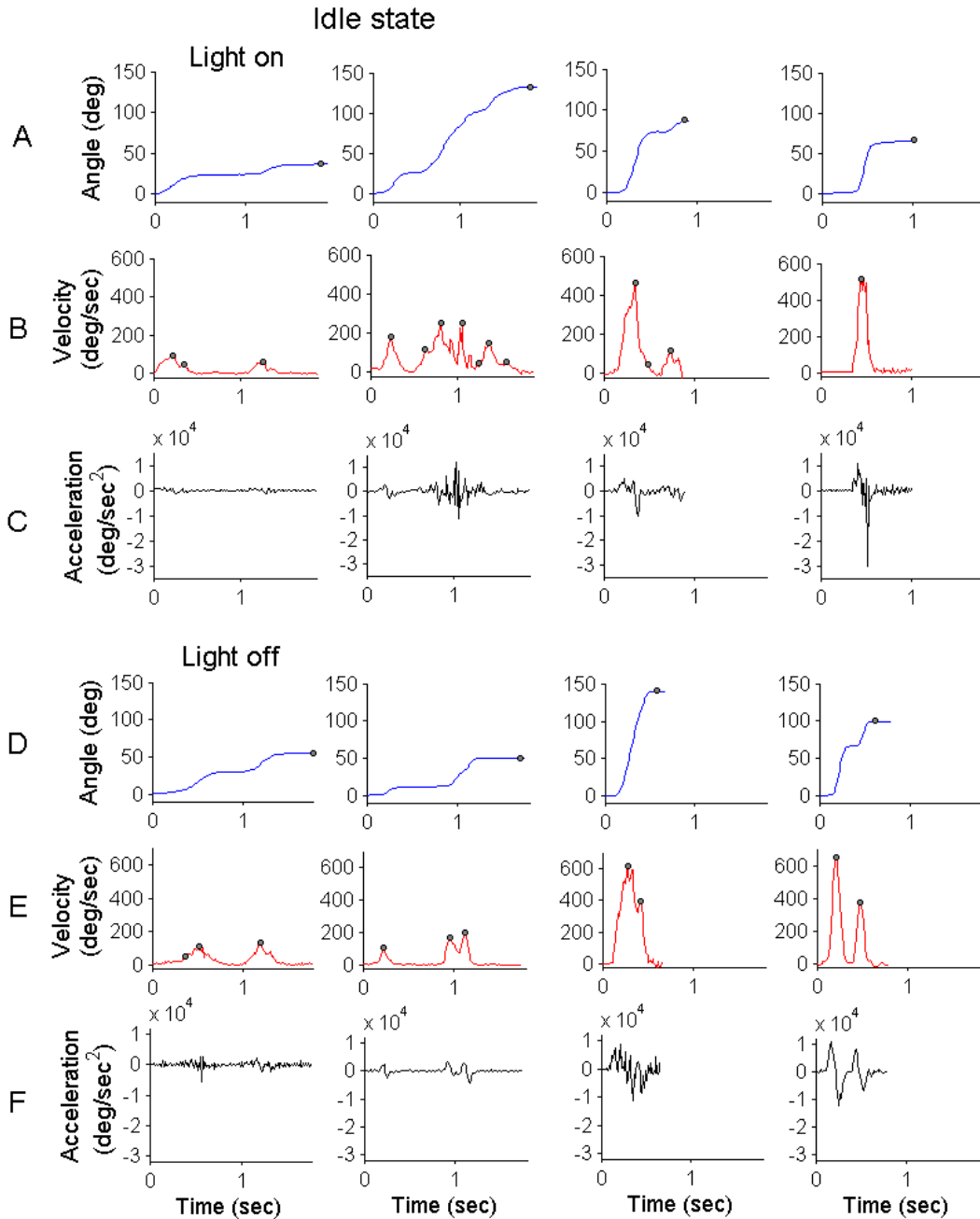
Figure 23: Example Results of Horizontal Head Movement Obtained in the Idle Condition When House Lights was On or Off

A. Four examples of horizontal rotation in angles obtained in light on condition, B. Velocity of horizontal rotation of those shown in A , C. Acceleration of horizontal rotations of those shown in A. , D. Four examples of horizontal rotation in angles obtained in light off condition, E. Velocity of horizontal rotation of those shown in D. F. Acceleration of horizontal rotations of those shown in D.

61

Figure 24: Example Results of Horizontal Head Movement Obtained in the Sound Stimulus Condition When House Lights was On or Off

A. Four examples of horizontal rotation in angles obtained in light on condition, B. Velocity of horizontal rotation of those shown in A , C. Acceleration of horizontal rotations of those shown in A. , D. Four examples of horizontal rotation in angles obtained in light off condition, E. Velocity of horizontal rotation of those shown in D. F. Acceleration of horizontal rotations of those shown in D.
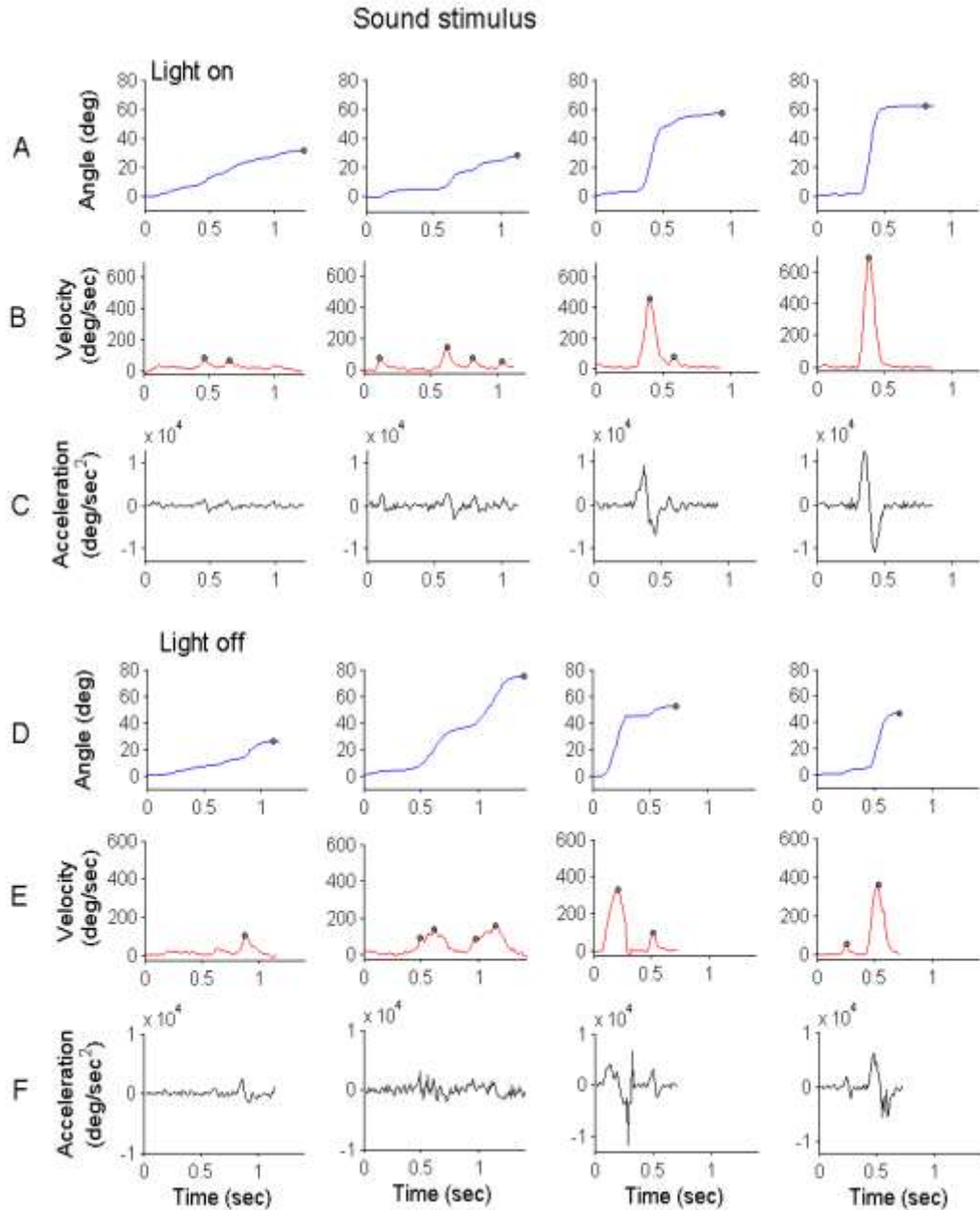
Figure 25: Example Results of Horizontal Head Movement Obtained in the Visual and Sound Stimulus Condition When House Lights was On or Off

A. Four examples of horizontal rotation in angles obtained in light on condition, B. Velocity of horizontal rotation of those shown in A , C. Acceleration of horizontal rotations of those shown in A. , D. Four examples of horizontal rotation in angles obtained in light off condition, E. Velocity of horizontal rotation of those shown in D. F. Acceleration of horizontal rotations of those shown in D.
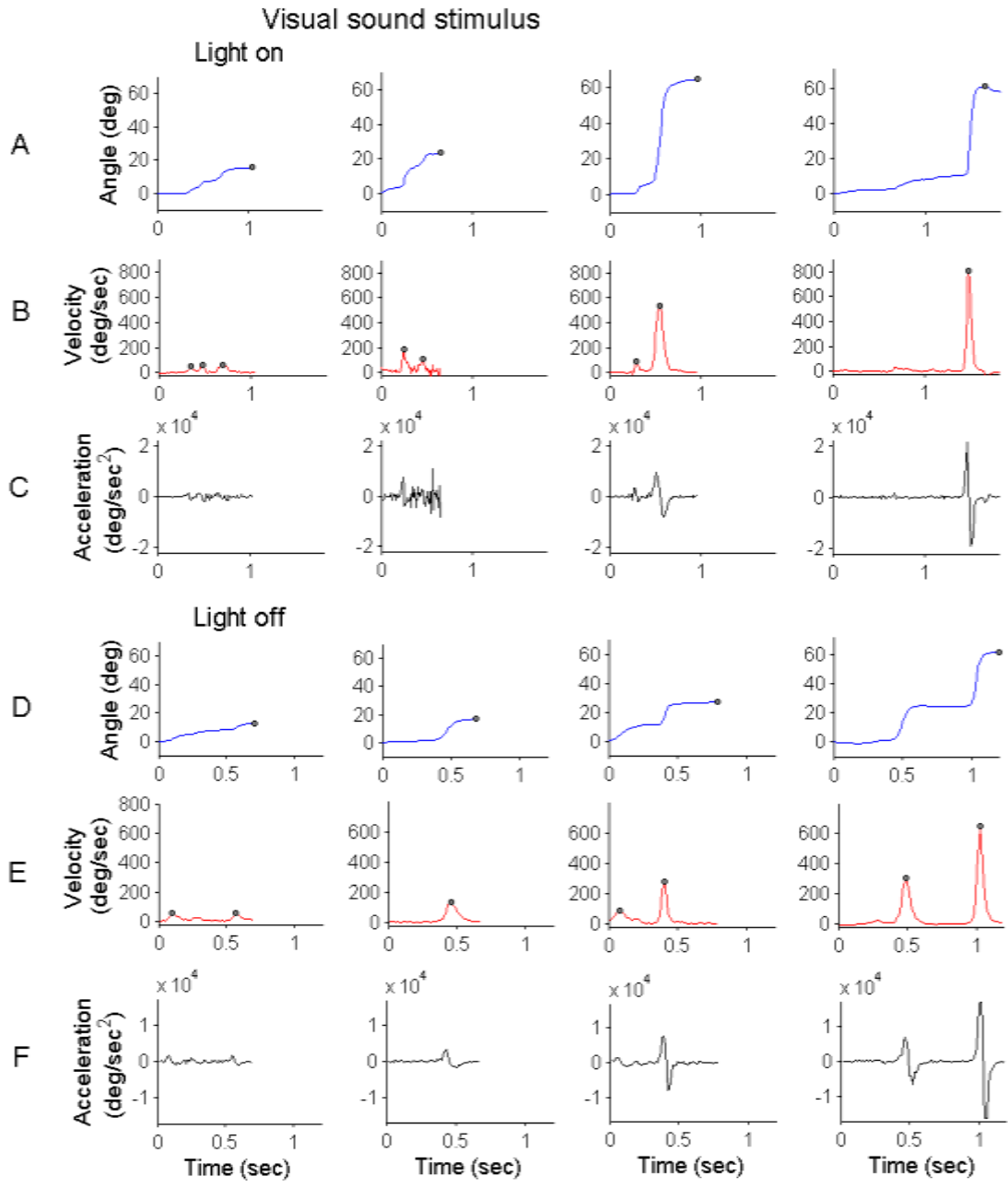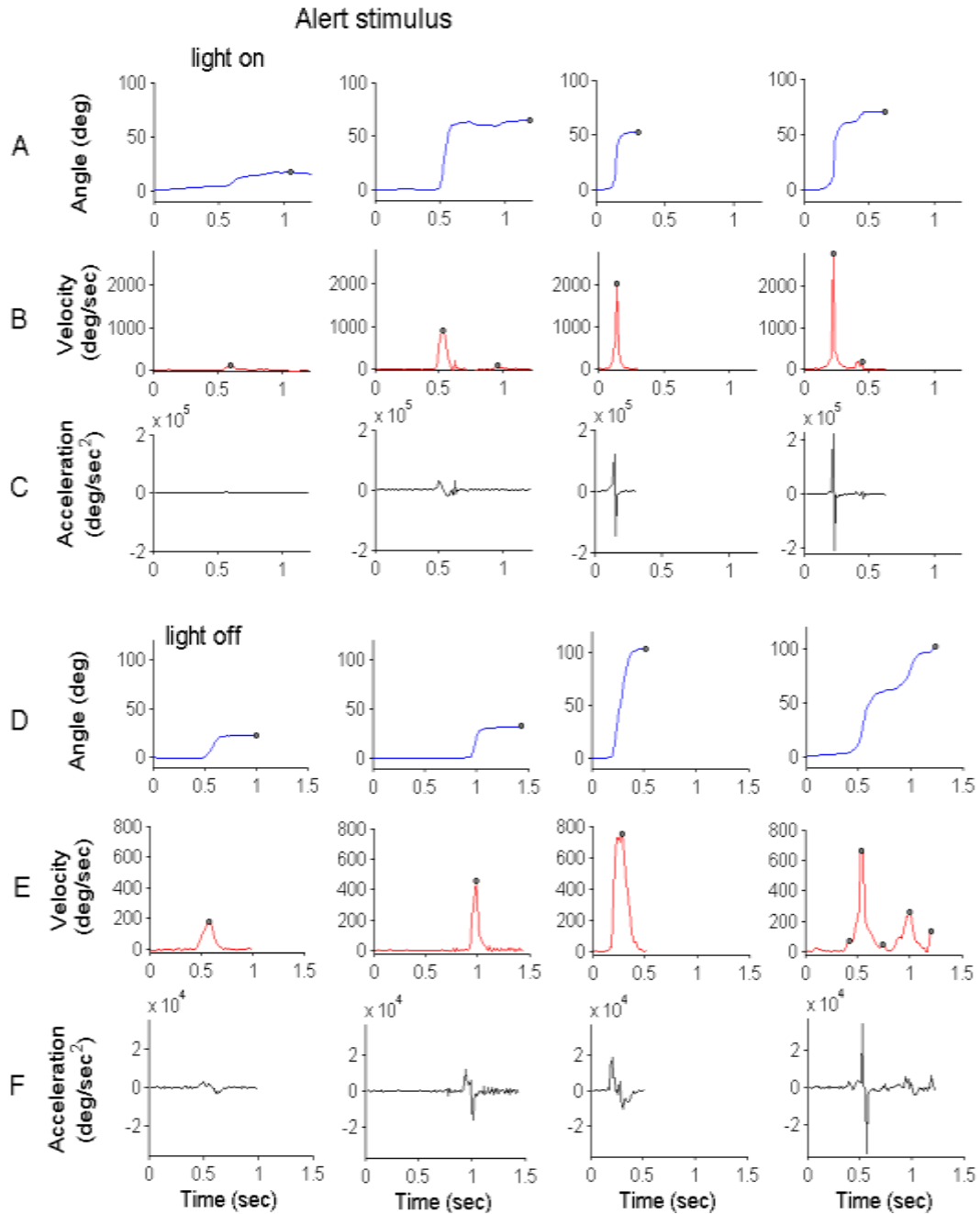
Figure 26: Example Results of Horizontal Head Movement Obtained in the Alert Stimulus Condition When House Lights was On or Off

A. Four examples of horizontal rotation in angles obtained in light on condition, B. Velocity of horizontal rotation of those shown in A , C. Acceleration of horizontal rotations of those shown in A. , D. Four examples of horizontal rotation in angles obtained in light off condition, E. Velocity of horizontal rotation of those shown in D. F. Acceleration of horizontal rotations of those shown in D.
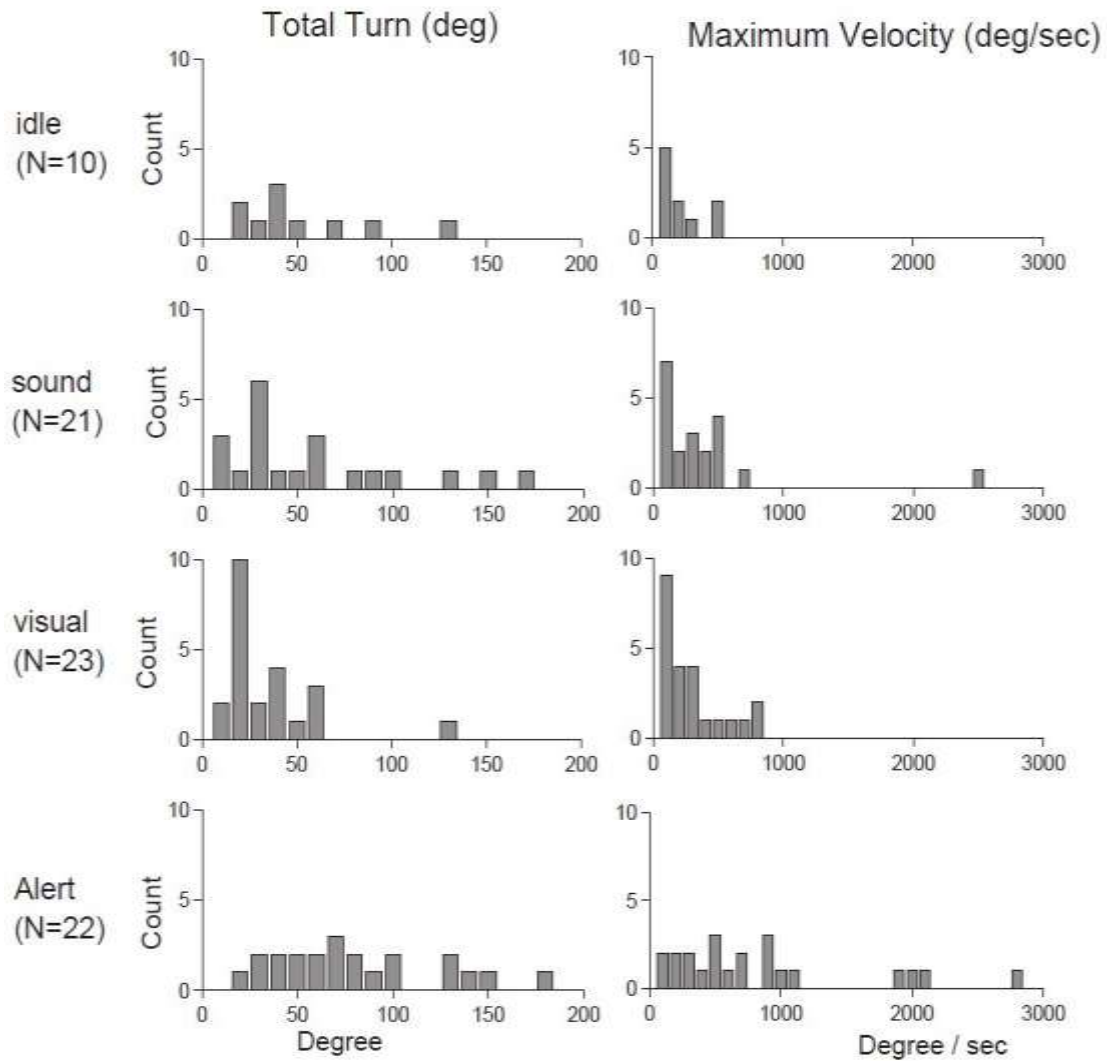
Figure 27 Histogram Plots for the Total Turning Angle of Horizontal Rotation and the Maximum Velocity for Different Stimulus Conditions with House Light On.

Left panels (A-D) show the total tuning angle and right panels (D-H) show the maximum velocity. The stimulus condition is marked on the top right corner on each panel. The total numbers of head turns in each condition are given in Table 5.
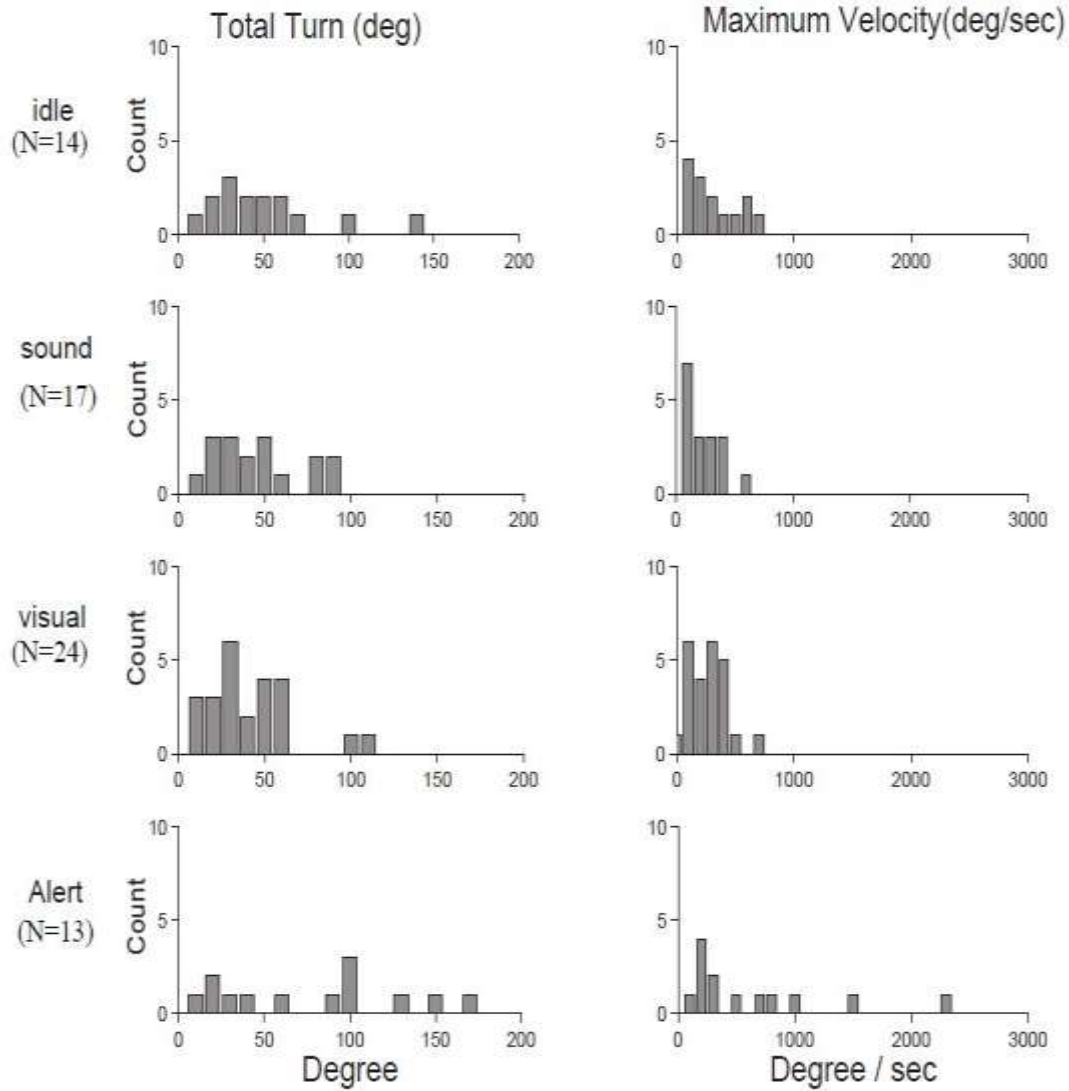
Figure 28 Histogram Plots for the Total Turning Angle of Horizontal Rotation and the Maximum Velocity for Different Stimulus Conditions with House Light Off.

Left panels (A-D) show the total tuning angle and right panels (D-H) show the maximum velocity. The stimulus condition is marked on the top right corner on each panel. The total numbers of head turns in each condition are given in Table 5.

Note that the bar plots do not have a Gaussian shape. From Figure 27 which has the

histograms for house light on condition, for panel A or idle condition, the histogram for

velocity is concentrated in the region <500 degrees per second and the maximal turn is spread from 0 to 150 degrees. In panel B or sound stimulus, the histogram for velocity is concentrated in the region <750 degrees per second and the maximal turn is spread from 0 to 170 degrees. In panel C or visual-sound stimulus, the histogram for velocity is concentrated in the region <1000 degrees per second and the maximal turn is spread from 0 to 70 degrees. In panel D or visual-sound stimulus, the histogram for velocity is distributed in both low and high velocity ranges, and highest bins are in the region of 2000 degrees per second and the maximal turn is spread from 0 to 200 degrees.

From Figure 28 which has the histograms for house light off condition, for panel A or idle condition, the histogram for velocity is concentrated in the region <750 degrees per second and the maximal turn is spread from 0 to 100 degrees. In panel B or sound stimulus, the histogram for velocity is concentrated in the region <500 degrees per second and the maximal turn is spread from 0 to 100 degrees. In panel C or visual-sound stimulus, the histogram for velocity is concentrated in the region <750 degrees per second and the maximal turn is spread from 0 to 100 degrees. In panel D or visual-sound stimulus, the histogram for velocity is distributed in both low and high velocity ranges, and highest bins are in the region of 1500 degrees per second and the maximal turn is spread from 0 to 175 degrees.

It can be seen that in general the bar plots are more concentrated towards the zero for coditions with house lights off. It can also be seen that maximum velocity occupies higher bins for the alert stimulus conditions.
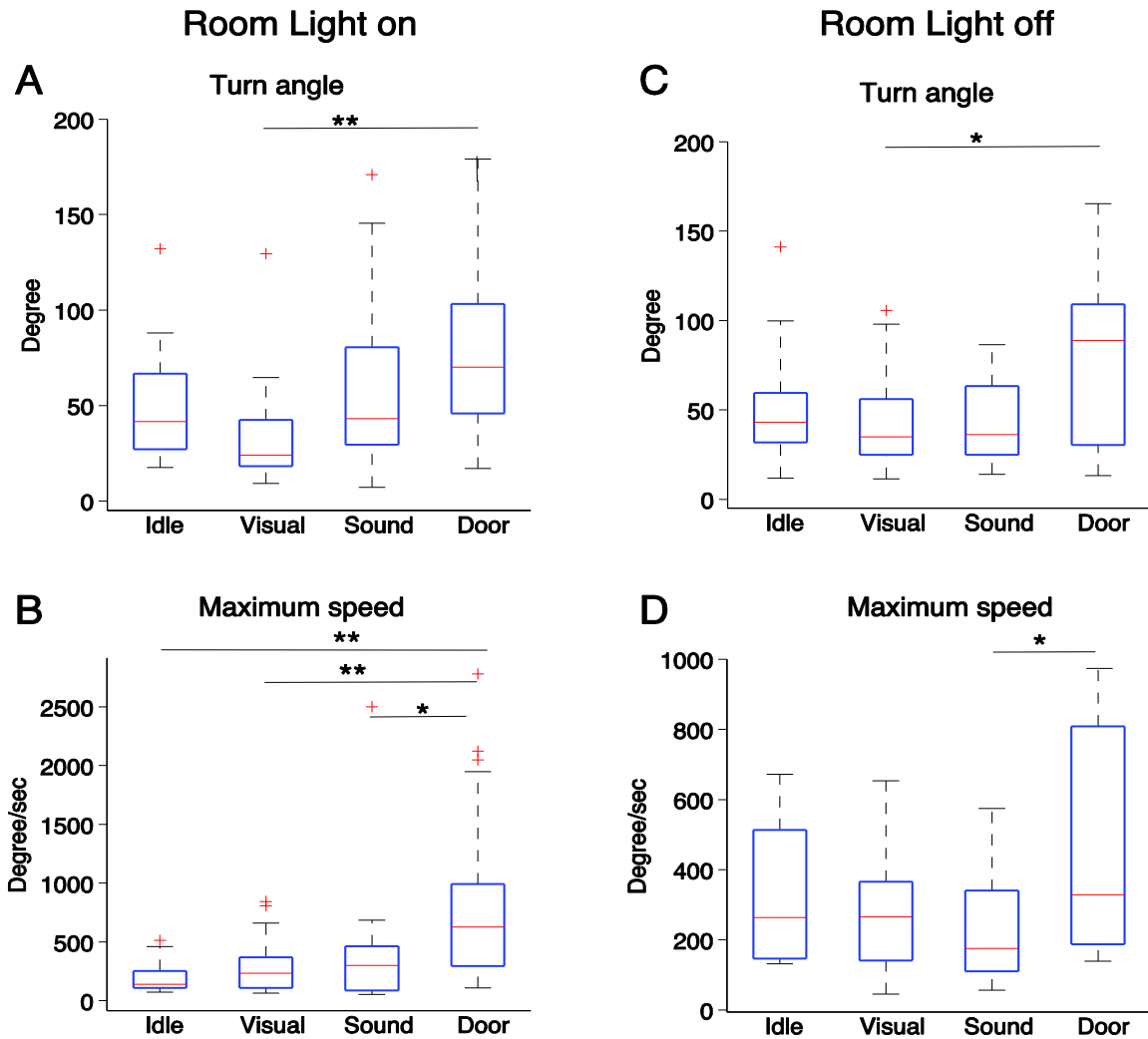
Figure 29: Statistical Analysis of the Maximal Head Turn and Maximal Velocity of Head Turn in the Horizontal Plane Tested with Four Different Stimuli.

A. The percentile of the maximal head turn angle when house light was on. B. The percentile of the maximal head turn velocity when house light was on. C. The percentile of the maximal head turn angle for house light off. D. The percentile of the maximal head turn velocity for house light off.   Pair-wise comparisons were made by Ranksum test. * indicates statistically significant difference with p<0.05, ** indicates statistically significant difference with p<0.01.

Figure 29 shows the statistical analysis of the results obtained from the different

conditions. The box-plots indicate the median with the 'red line', the 25 percentile and 75

percentile data points with the edges of the blue box. The outliers are indicated with the red '+' symbols. The 'ranksum' function in Matlab was used for analysis and the resulting p-value was the criterion for significant difference. Figure 29 shows the pair-wise comparisons that were made by Ranksum test and * indicates statistically significant difference with $p<0.05$ whereas ** indicates statistically significant difference with $p<0.01$. The median values for each maximum velocity distribution in order of idle, sound stimulus, visual stimulus and alert stimulus for house light on in degrees per second are: 139.56, 300.33, 232.7 and 627.8. The median values for each maximum turn angle distribution in order of idle, sound stimulus, visual stimulus and alert stimulus for house light on in degrees are: 41.53, 43.03, 23.07 and 70.08. The median values for each maximum velocity distribution in order of idle, sound stimulus, visual stimulus and alert stimulus for house light off in degrees per second are: 263.25, 175.23, 265.79 and 328.08. The median values for each maximum turn angle distribution in order of idle, sound stimulus, visual stimulus and alert stimulus for house light off in degrees are: 43.15, 36.23, 34.97 and 88.8. The medians for the alert stimulus are higher than the other stimuli.

Panel A indicates that for house light on condition, there is a statistically significant difference between the distributions for maximal head turn for visual-sound stimulus and the alert stimulus. Panel C shows the same with house lights off. From Fig. 29B it can be seen that the significant difference is highest for the alert stimulus condition with any of the other stimulus conditions. Panel D indicates that for house light off condition, there is a statistically significant difference between the distributions for maximal head turn for sound stimulus and the alert stimulus.

Thus it can be seen that there is a relation between the maximum velocity and the type of

stimulus being presented.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

The current thesis developed a mechanism to detect and analyze the head movements of marmoset monkeys. Although there are a number of existing methods and commercial products for motion tracking, they pose different challenges to be implemented for this current research with respect to cost, sensor size and types. This necessitated custom methods for tracking fast head movement of marmoset monkeys. Our method used infra-red camera with a frame rate of 90 Hz was used to capture the head movement of monkeys. To assist in the signal detection, two circular markers were attached on top of the monkey head. For data analysis, blob detection technique was used to obtain locations of markers. A tracking method using Kalman filter was further added to the blob detection to obtain the marker locations even when markers were temporarily out of the field of the view of the camera. For validation, results were compared with other techniques such as optical flow and SURF using constant-speed data from a DC motor. Root mean square errors and time duration analyses showed that the blob detection algorithm we developed was better suited for the purpose of this research. The horizontal and vertical rotation angles, the corresponding velocity and acceleration were calculated based on the blob detection method.

In result analysis, statistical tests were performed on horizontal rotation angle and velocity data extracted from the segmented image files. Data were grouped based on four experimental conditions: (1) idle state with no auditory and visual stimuli being presented, (2) sound stimulus, (3) visual and sound stimulus, and (4) alert signal by opening and closing the chamber door. The data for each condition was further divided

71

into two groups according to whether house lights were on or off. The results for different conditions were compared with the idle state and with each other. It was found that the results obtained with the alert signal (door opening/closing) had considerably higher velocities than the results obtained under other conditions. In addition, the velocities were significantly higher when the house light was turned on than the house light was turned off. These results suggest that the speed of head movement in marmoset monkeys is correlated with the stimulus conditions. Natural signals, such as opening the door of the chamber represents ecologically relevant stimuli for marmoset monkeys, as they tend to suggest a test session is over, and thus evoke stronger and faster behavior output in marmosets.

For future research, some of the challenges facing this research could be addressed.

One challenge is that the current method using one camera cannot reliably obtain the information of vertical rotation. This could be overcome by adding additional cameras to the setup to resolve the directions of vertical rotations.

The second challenge is to obtain accurate ground truth data for validation. Since the ground truth in the current project was obtained by manually marking the location of markers on images, it may be prone to human errors. In the future, an accelerometer system could be used to continuously monitor the head movement in both horizontal and vertical planes. This, however, would require a monkey subject with a head implant to allow access to monkey head and to provide support for mounting the recording devices.

The third challenge is to obtain sufficient numbers of behavior responses. As mentioned earlier, since the animals used in the current research were not behaviorally trained, they

were not required to respond to stimulus. As a result, a large number of the trials had to be collected in order to obtain trials with head movements. Adaptation and fatigue were observed in experiments. To overcome this problem, future research could test head turning behaviors using trained monkeys, thus reducing the trial number. For a trained monkey, the rotation velocities of head movement are likely to differ from those obtained by this research, which focused on the reflexive responses of head. It would be of great interest to explore such differences in the future.

Fourth, the head motion was analyzed off-line and the stimuli were presented in random orders in the current approach. Further research may benefit from a real-time implementation in which head motion is analyzed as the stimulus is presented and. The real-time analysis will cast a high demand on the computing powers of the devices and algorithms. Currently, many new graphic tools are available for increasing the speed of video and image processing. Graphics processing Units (GPU computing) would enhance the speed of execution of a program to a great extent.

Finally, there is a need to understand the effects of moving stimuli as opposed to the stationary stimuli, on head movement. Natural environment contains rich and dynamic sensory stimulation, as opposed to the stationary ones used in the current research. In natural settings, animals as well as their predators can move around. Thus it would be more natural to study the movements of the monkeys when their body is not constrained. For such research, portable devices such as accelerometers and wireless MEMS recording chips may greatly advance the capacity of this research to quantify the natural behaviors of marmoset monkeys.

In summary, the methods developed in this thesis could serve as a stepping stone to future research in understanding the head movement behaviors of marmoset monkeys. The results obtained in this thesis demonstrate the importance of stimulus features in evoking head rotation behaviors. This information is important for unveiling the role of self-elicit movement in improving object recognition and localization in natural environments.

There could be research on the effects of moving stimuli as opposed to the stationary stimuli used in this research, since in a natural environment, animals would be facing such stimuli. Also since in natural conditions the body of the animal is not in a fixed location as in this experiment, it would be more natural to study the movements of the monkeys in the colony cages. Even for this kind of research, accelerometers could play an important role.

For all the mentioned future research, the material presented in this thesis could be a stepping stone. Revelations could be made that validate the results of this thesis, or the results could be used as a validation mechanism for other methods.

References

Bay, H., Tuytelaars, T., & Gool, L. V. (2006). *SURF: Speeded up robust features. Proceedings of the Ninth European Conference on Computer Vision.*

Bendor, D., & Wang, X. (2005). *The neuronal representation of pitch in primate auditory cortex. Nature, 436*(0028-0836), 1161-1165.

Blauert, J. (Ed.). (1997). *Spatial hearing: The psychophysics of human sound localization.* Cambridge, MA.: MIT Press.

Bradski, G. R., Kaehler, A., & Safari Books Online. (2008). *Learning OpenCV* (1st ed.). Sebastopol, Calif.: O'Reilly.

Brimijoin, W. O., Boyd, A. W., & Akeroyd, M. A. (2013). *The contribution of head movement to the externalization and internalization of sounds. Public Library of Science, 8*(12), e83068.

Burman, K. J., Reser, D. H., Yu, H. H., & Rosa, M. G. (2010). *Cortical input to the frontal pole of the marmoset monkey. Cereb Cortex,* **21**, 1712-1737.

Cloherty, S. L., Mustari, M. J., Rosa, M. G., & Ibbotson, M. R. (2010). *Effects of saccades on visual processing in primate MSTd. Vision Res, 50*(24), 2683.

Common marmoset care. Retrieved July,30, 2014, from http://www.marmosetcare.com/

Epple, G. (1968). *Comparative studies on vocalization in marmoset monkeys (hapalidae). Folia Primatol (Basel),* **8**, 1-40.

Fleet, D. J., & Weiss, Y. (2005). *Optical flow estimation. Mathematical Models in Computer Vision: The Handbook.,* 239.

Geometric transformations. (2002). Retrieved july/30, 2014, from www.cs.mtu.edu/~shene/COURSES/cs3621/NOTES/geometry/geo-tran.html

Guitton, D., & Volle, M. (1987). *Gaze control in humans: Eye-head coordination during orienting movements to targets within and beyond the oculomotor range. Journal of Neurophysiology, 5, No 3.*

IGS-190 motion capture system. (2010). Retrieved July, 30, 2014, from http://www.metamotion.com/gypsy/IGS-190-mobile-motion-capture-system.html

Marmoset-vocalizations. (2013). Retrieved July, 30, 2014, from http://www.marmosetmom.com/Marmoset-Vocalizations.html

Middlebrooks, J. C., & Green, D. M. *Sound localization by human listeners. Annu Rev Psychol, 42*, 135-159.

Murphy-Chutorian, E., & Trivedi, M. M. (2009). *Head pose estimation in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31,* No. 4, April 2009. 10.1109/TPAMI.2008.106

Natural point TrackIR, Retrieved from: https://www.naturalpoint.com/trackir/.

Osmanski, M. S., & Wang, X. (2011). Measurement of absolute auditory thresholds in the common marmoset (callithrix jacchus). *Hear Res, 277*, 127-133.

Otsu, N. (1979). *A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9*, No. 1, pp. 62-66, 1979.

Primate Info Net. *Behavior of the common marmoset.* Retrieved July, 30, 2015, from http://pin.primate.wisc.edu/callicam/calbeh.html.

Seiden, H. R. (1957). *Auditory acuity of the marmoset monkey (hapale jacchus).* Princeton University). *PhD Thesis,*

Seitz, S. (2009). *Optical flow equation,* Retrieved from http://www.cs.umd.edu/~djacobs/CMSC426/OpticalFlow.pdf

Slee, S. J., & Young, E. D. (2010). *Sound localization cues in the marmoset monkey. Hear Res, 260*, 96-108.

Spoor, F., Garland, T., Krovitz, G., Ryan, T. M., Silcox, M. T., & Walker, A. (2007). *The primate semicircular canal system and locomotion. Pnas,* vol. 104, No. 26,*104.*

Thurlow, R. W., Mangels, J. W., & Runge, P. S. (1967). *Head movements during sound localization. The Journal of the Acoustical Society of America.*

Visage Technologies. *Visage technologies tracking and animation.* Retrieved July,30, 2014, from http://www.visagetechnologies.com/products/visagesdk/headtrack/

Wallach, H. (1940). *The role of head movements and vestibular and visual cues in sound localization*. *Journal of Experimental Psychology, 27,* No. 4.

What is wii U? (nd). Retrieved July,30, 2014, from http://www.nintendo.com/wiiu/what-is-wiiu

what-when-how. (2014). *BLOB analysis (introduction to video and image processing) part 1*.http://what-when-how.com/introduction-to-video-and-image-processing/blob-analysis-introduction-to-video-and-image-processing-part-1/

Wightman, F. L., & Kistler, D. J. (1999). *Resolution of front-back ambiguity in spatial hearing by listener and source movement*. *The Journal of the Acoustical Society of America., 105*, 2841.

Young, D. (2010,). Affine optic flow. Message posted to http://www.mathworks.com/matlabcentral/fileexchange/27093-affine-optic-flow/content/affine_flow.m

Zangemeister, W. H., & Lawrence, S. (1981). *Active head rotations and eye-head coordination*. *New York Academy of Sciences,* 0077-8929/81/0374-0540.