Project Proposal: Team 29

Gaze and Speech Tracking with Haptic Feedback for the Visually Impaired

Hannah Burke – Zeen Luan – Nicolas Chapman

Motivation and Target Customer

Visually impaired (VI) people rely heavily on their sense of touch, smell and sound to interact with the world around them, and do so remarkably well with the aid of only basic tools such as canes. However, the cognitive load associated with completing seemingly simply tasks is often extreme, especially in noisy and crowded environments. The development of wearable robotic systems has been largely targeted at reducing this cognitive load. Such robotic systems generally have two components; a sensing mechanism and a feedback mechanism. The sensing components used in existing devices include ultrasonic sensors for obstacle recognition [4], GPS for navigation [1] and cameras for object classification [2][3]. Spoken feedback is traditionally used to articulate information to users, but tactile feedback is becoming more popular as it is less intrusive on the users' other senses, and is more effective in noisy environments [1].

Developments in computer vision and artificial intelligence have been particularly important for the VI, enabling them to engage with their environments, and particularly other people, in ways that were previously impossible. For example, prototypes have been developed that convey facial expressions to users via wearable vibrotactile devices [3]. Another interesting application of such technology is the tracking and communicating of gaze to the VI. The E-Gaze Glasses offered a conceptual design for tracking the gaze of a single speaker, and communicating to the user if they are being looked at [2]. Gaze is an extremely important conveyor of attention in conversation, and without this information, it can be difficult to discern who is the intended recipient of speech.

Our product will build on the design of the E-Gaze Glasses to convey to the user via haptic feedback the location of people in their field of vision, if these people are looking at them and if they are talking. It will work in dynamic environments where people are moving around, into and out of the field of vision. It will also work with more than one person present in the field of vision. We hope that such a product will greatly reduce the cognitive load of VI people when they engage in conversation by clearly communicating to them if they are the intended recipient of speech.

Functions and Specifications

As previously mentioned, our proposed product will track people in the camera field of view, compute if these people are looking at the camera and if they are talking and communicate this information to the user via a vibrotactile device. The sensing component of the product will consist of a servo-controlled head mounted camera with pan and tilt capability. The feedback component will consist of a five vibrating motor disks arrange evenly around the head; one in the middle of the forehead, one on each corner of the forehead, and one on each side of the head.

Function 1: Camera Tracking of People in Field of Vision

When there is no person in the field of vision, the camera pans left and right repeatedly to cover the full field of vision. The field of vision is defined as 90 degrees left and right of centre, as it is roughly within this range that gaze tracking is used by sited people. When a person is recognised using computer vision tools, the camera tracks the location of their face while they remain in the field of vision.

Function 2: Communication of Gaze and Speech

If the person being tracked is found to be looking at the camera, they are deemed a 'potential speaker' and the haptic motor denoting their direction relative to the user is buzzed softly until eye contact is broken for a considerable period. If the person looking at the camera is speaking, they are deemed a 'speaker' and the haptic motor will be buzzed with greater intensity.

Function 3: Tracking with Multiple People

The potential for multiple people to be in a single frame complicates the tracking process. To solve this, the device priorities tracking of 'speakers' over 'potential speakers', and 'potential speakers' over 'non-speakers'. That is:

- A person looking at the camera and speaking (speaker) will be tracked over those only looking at the camera (potential speaker) and those doing neither (non-speaker).
- A person looking at the camera (potential speaker) will be tracked over those not looking at the camera (non-speaker).

If there are two people in a frame of the priority to be tracked, the closest person to the user will be selected for tracking based on the size of their face.

Product Highlights

A substantial appeal of our product is that it will be inexpensive. It will be a fraction of the price of existing computer vision navigation aids; for example, Vision Australia's product MiniGuide Ultrasonic Echo-location Detector which is $741 [4]. Our estimated budget for a single prototype is $241.41 (see below for details) and so could be manufactured and sold with a sizable mark up at a much lower price than competing products.

We believe our focus on reducing cognitive overload will also be a marketable attribute of our product. Our competitors offer a range of handheld products which the user must trigger and various complex designs with a wider range of features. However, our product will be hands free and autonomously run, allowing the user to concentrate on their interactions being assured that aid will be given to them when required. Moreover, our targeted focus on gaze and speech recognition combined with the simple nature of haptic feedback ensures that the user will be able to interpret the signal intuitively (especially after ongoing use).

Lastly, our devices ability to deal with multiple people in the field of vision further means it is effective in crowded and noisy environments, solving a key limitation of the conceptualised E-Gaze Glasses.

Market Size

In Australia, there are 575,000 blind and visually impaired people [5], proving there is a large market of potential buyers. In 2018, Vision Australia alone made $8.3 million from the sales of aids [6], and given there are many other retailers of computerized aids, it is apparent that this is a sizeable industry. While the product is clearly an appropriate personal aid for individuals, we also intend on marketing the device to organisations. This product will be particularly useful when it is necessary that communication between several people is seamless and effective. Thus, we believe it should be on hand in classrooms and meeting spaces, even without being designated for a specific user, to ensure these spaces are disability friendly. Consequently, schools and offices will provide a considerable portion of the market.

Resources Required and Budget Plan

Table 1: Resources required and costs for all components of prototype

	Resource	**Purpose**	**Quantity**	**Cost**
Provided by Monash	Raspberry Pi 3	Core microprocessor used in the project. It will handle the image capture, run the required computer vision algorithms and communicate with the Arduino.	1	$70.00
	Raspberry Camera	Used to capture images of the surroundings.	1	$38.95
	5V DC Power Plug	Powers the RPI.	1	$21.95
	Solderless Breadboard	Motor control circuit.	1	$6.58
	Wires	Motor control circuit.	~ 10	$4.40
	Resistors	Motor control circuit.	~ 10	$1
	Pan Tilt Camera Bracket	Control the orientation of the camera in tracking process.	1	$12.87
	Micro Servo Motor	Control the orientation of the camera in tracking process.	2	$15.80
	3D Printed Head Mount	Comfortably and safely mount the system on the user's head.	1	$5 (max)
		Total		**$176.55**
Purchased	Vibrating Mini Motor Disks	The haptic actuators chosen to provide feedback to the user.	5	$22.00
	Arduino	Communicates with the RPI to control the output actuators.	1	$30.00
	PN2222 Transistor	Motor control circuit.	10	$4.40
	Velcro Strip	Attach components to head mount.	50cm	$5
	M3 Screws	Attach components to head mount.	~ 10	$2
	Resistors (330ohm)	Motor control circuit.	20	$1.46
	Computer with relevant software	Required for design and programming of system and potentially cloud computing.	1	N/A
		Total		**$64.86**

Project Time Plan

We have chosen to split our parts for the first 10 weeks into parallel components, in case the lockdown results in us not being able to exchange equipment. The Arduino, motor controllers and haptics can be in one location while the raspberry pi and camera can be in another, to ensure parallel development of the project.

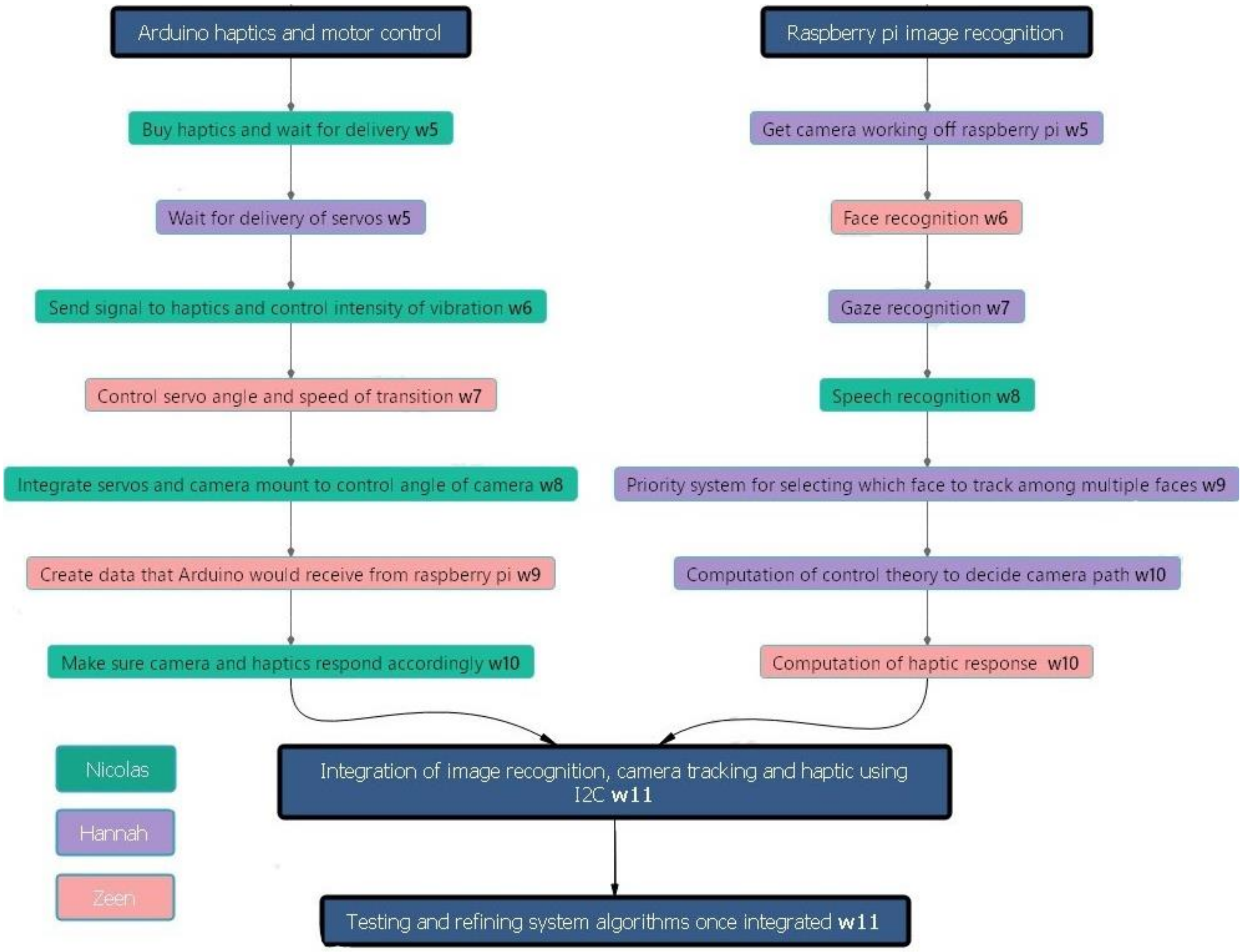

Figure 1: Project time plan flow chart

Responsibility Matrix

Table 2: Responsibility matrix for key tasks

Week	Task	Nicolas	Hannah	Zeen
5	**Buy haptics and wait for deliveries**	P	S	S
	Get camera working off raspberry pi	S	P	
6	**Send signal to haptics and control intensity of vibration**	P	S	
	Face recognition	S	S	P
7	**Control servo angle and speed of transition**	S		P
	Gaze recognition	S	P	S
8	**Integrate servos and camera mount to control angle of camera**	P	S	
	Speech recognition	P	S	S
9	**Create data that Arduino would receive from raspberry pi**	S	S	P
	Priority system for selecting which face to track among multiple faces	S	P	S
10	**Make sure camera and haptics respond accordingly**	P		S
	Computation of control theory to decide camera path	S	P	
	Computation of haptic response		S	P
11	**Integration**	P	S	S
11	**Testing and refining**	S	P	S

Key:

- P: primary
- S: support
- Arduino haptics and motor control tasks
- Raspberry pi image recognition tasks

References

[1] Kammoun, Slim & Jouffrais, Christophe & Guerreiro, Tiago & Nicolau, Hugo & Jorge, Joaquim. (2012). Guiding Blind People with Haptic Feedback.

[2] Qiu, Shi & Anas, Siti & Osawa, Hirotaka & Rauterberg, Matthias & Hu, Jun. (2016). E-Gaze Glasses: Simulating Natural Gazes for Blind People. 563-569. 10.1145/2839462.2856518.

[3] Buimer HP, Bittner M, Kostelijk T, van der Geest TM, Nemri A, van Wezel RJ, Zhao Y. Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device. PLoS One. 2018;13(3):e0194737. doi: 10.1371/journal.pone.0194737.

[4] Vision Australia: Everyday Technologies. https://shop.visionaustralia.org/shop/in-department/departments/everyday-technology#

[5] https://everyaustraliancounts.com.au/research-reports/a-snapshot-of-blindness-and-low-vision-in-australia/#:~:text=It%20is%20estimated%20there%20are,66%2C000%20people%20who%20are%20blind.

[6] https://www.visionaustralia.org/sites/default/files/8439-VA-Financial%20Report-2018-WEB.pdf