VISTA – CVR
Virtual Vision Futures

June 14 – 17, 2021

Conference Program

Centre for Vision Research

YORK UNIVERSITÉ UNIVERSITY

VISTA
VISION: SCIENCE TO APPLICATIONS

VISTA – CVR
Virtual Vision Futures

Table of Contents

Schedule at-a-glance 3

Monday, June 14th 3

Tuesday, June 15th 5

Wednesday, June 16th 7

Thursday, June 17th 9

Presentation abstracts 11

Workshops 77

Non-academic jobs 77

Academic jobs 78

Impactful scientific writing and knowledge dissemination 79

VISTA – CVR
Virtual Vision Futures

Schedule at-a-glance

Monday, June 14th	**Early Session 8:30 – 9:20am**	**IRTG Session #1** Chair: Dr. Denise Henriques **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Dr. Denise Henriques	Welcome remarks
		Elef Schellen and Edward Ody	Introduction
		Raphael Gastrock	Skill acquisition versus adaptation: Using a mirror reversal task to investigate de novo learning and distinguish it from motor adaptation
	Breakout Sessions 9:30 – 11:10am	**IRTG Session #2** Chair: Mahboubeh Habibi **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Lisa Rosenblum	Influence of tactile flow on visual heading perception
		Elef Schellen	Neural markers of size perception in dynamic environments
		Laura Mikula	Interpreting and integrating vision for adapting to dynamic environments
		Edward Ody	Sensory suppression of self-generated visual and auditory action consequences.
		Hamid Ramezanpour	Temporal cortex participates in attention control
		9:20am	**VVF Opening Remarks: Dr. Robert Allison** **Zoom link: https://tinyurl.com/VirtualVisionFutures2**
		VVF Session #1 Chair: Jason Yu **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Gaeun Son	Scene wheels: Measuring perception and memory of real-world scenes with a continuous stimulus space
		Katerina Andrinopoulos	Sex differences in visuospatial mental rotation persist under 3D VR conditions
		Jason Yu	Wavelet flow: Fast training of high-resolution normalizing flows
		Jingmin Zhou	Multi-label video categorization
		Katherine Newman	Understanding the influence of a brain derived neurotrophic factor gene polymorphism on brain network integrity and non-invasive brain stimulation techniques

VISTA – CVR
Virtual Vision Futures

Monday, June 14th - continued	**Breakout Session 11:20am – 1pm**	**VVF Session #2** Chair: Jennifer Ruttle **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Ala Salehi	Unsupervised data mining from videos for improved depth estimation in driving datasets
		Assel Al-Bayati	Frequent cannabis-use has no long-term effects on visuo-motor and visuo-cognitive performance
		Caroline Giuricich	Target-distractor competition modulates saccade trajectories
		Jennifer Ruttle	Implicit components of motor learning saturate faster with full control of movement and visual feedback
		Sebastian D'Amario	The influence of demographics, lifestyle and external events on cognitive and motor tasks
		VVF Session #3 Chair: Gaelle Luabeya **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Ayoob Shahmoradi	Representation as representation-as
		Bianca Baltaretu	A cortical network for trans saccadic perception of object features: An fMRI paradigm
		Huiqin Chen	Perceptual hysteresis in the categorization of complex scenes
		Gaelle Luabeya	Neural mechanism of integration of location and orientation in preparation for grasp
		Tasfia Ahsan	Perceived depth modulates perceptual resolution
	1 - 2pm	**Lunch @Gather Town** **Link: https://gather.town/i/IeCDDqJ6; Passcode: 3291** Chat with presenters, hang out with colleagues and friends	
	2 - 4pm	**Workshop #1 – Panel discussion on non-academic jobs (see p.77)** **A discussion about applying for non-academic jobs across disciplines, with: Dr. Raghavender Sahdev** (NuPort Robotics Inc); **Dr. Caitlin Mullin** (VISTA @ York University); **Dr. Soo Min Kang** (Samsung AI); **Dr. Lindsey Fraser** (Vpixx Inc.); **Dr. Carolyn Steele** (Career Development coordinator, York University) **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	

Tuesday, June 15th	**Early Session 8:30 – 9:20am**	**IRTG Session #3** Chair: Moritz Schubert **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Christina Schmitter	Discrete vs. continuous action feedback: Commonalities and differences in predictive neural processing
		Ben Cuthbert	Visual working memory models do not generalize to the whole-report task
		Ilja Wagner	Humans trade-off set-size and discrimination difficulty in a combined visual search and perceptual discrimination task
	Breakout Sessions 9:30 – 11:10am	**IRTG Session #4** Chair: Jonathan Coutinho **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Brandon Caie	Predictive Gating of Evidence Accumulation
		Frieder Tom Hartmann	Hand configurations and contact regions of natural human grasps
		Mahboubeh Habibi	A free viewing paradigm to identify eye movement disturbances in alpha-synucleinopathies versus tauopathy
		Elena Fuehrer	Sensorimotor predictions lead to sensation-specific tactile suppression
		Johannes Keck	Spatiotemporal features of emotional body language in social interactions
		VVF Session #4 Chair: TBD **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Paria Mehrani	Learning a sparse representation in V4 cells reveals shape encoding mechanisms in the brain
		Rachel Moreau	Differential processing of reflection and rotation symmetries in visual textures
		Rebecca Keller	(Endogenous) perceptual states are conceptual
		Sidharth Gupta	Implementing and integrating contour completion using Perceptual Grouping
		Krista Mitchnick	Damage to the hippocampal dentate gyrus impairs the perceptual discrimination of complex, novel objects

VISTA – CVR
Virtual Vision Futures

Tuesday, June 15th - continued	**Breakout Session 11:20am – 1pm**	**VVF Session #5** Chair: Alica Rogojin **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Alica Rogojin	White matter microstructure changes are associated with declines in cognitive-motor task performance in older adults with a genetic (APOE e4) risk for Alzheimer's disease
		Mylann Guevara	The neural underpinnings of mental attention capacity in healthy young adults
		Rebecca Barnstaple	Visuomotor integration in dance-based learning: Applications for clinical populations
		Sara Pishdadian	Mnemonic discrimination and spatial abilities in healthy aging and subjective cognitive decline
		Thomas Chen	Computer vision for social good: adapting to the changing climate and extreme weather events
		VVF Session #6 Chair: Björn Jörges **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Björn Jörges	Visually simulated self-motion biases the perception of object motion and makes it less precise
		Brittney Hartle	Scaling stereoscopic depth through reaching
		Ewen Lavoie	Embodiment, visual attention, and movement in real and virtual worlds
		Fengbo Lan	Joint demosaicking / rectification of fisheye camera images using multi-color graph laplacian regularization
		Domenic Au	The impact of binocular capture on monocular content in augmented reality
	1 – 2pm	**Lunch @Gather Town** **Link: https://gather.town/i/leCDDqJ6; Passcode: 3291** Chat with presenters, hang out with colleagues and friends	
	2 – 4pm	**Workshop #2: Panel discussion on applying for academic jobs (see p.78)** **Our panelists: Dr. Kevin Lande** (Philosophy); **Dr. Laurence Harris** (Psychology); **Dr. Gene Cheung** (EECS)**; Dr. Caitlin Fisher** (Arts); **Dr. Shital Desai** (Design) **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	

VISTA – CVR
Virtual Vision Futures

Wednesday, June 16th	**Early Session 8:30-9:20am**	**IRTG Session #5** Chair: Lisa Rosenblum **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Jakob Schwenk	Processing of temporal information in marmoset primary visual cortex
		Jonathan Coutinho	Luminance evoked pupil response dynamics
		Renate Reisenegger	Neurophysiological correlates of self-motion processing and navigation
	Breakout Sessions 9:30 – 11:10am	**IRTG Session #6** Chair: Laura Mikula **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Matt Laporte	Robust versus optimal control of movements by neural networks
		Nedim Goktepe	Understanding the temporal dynamics and informational content of foveal feedback in peripheral vision.
		Lucie Preißler	Emotion perception of interactive body movements in preschoolers
		Kieran Hussey	Familiar size reliably affects size and distance perception in high-resolution virtual reality
		Lina Musa	Explicit attention to allocentric visual landmarks improves memory-guided reaching
		VVF Session #7 Chair: Tenzin Chosang **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Hossein Pourmodheji	Multiple pedestrian tracking by a mask R-CNN with post processing and an adaptive information-driven motion particle filter model
		Harpreet Saini	Color modulates feature integration
		Sarah Park	Will the colavita effect persist in online testing?
		Tiina Rosenqvist	What are colour visual systems doing?
		Tenzin Chosang	Evaluating the perceptual efficiency of shape representation models
	Breakout Session 11:20am – 1pm	**VVF Session #8** Chair: Sarah Vollmer **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Lee Williams	I am your ghost
		Jonathan Tong	Curvature detection and discrimination thresholds for parabolic surfaces depend on the direction of curvature
		Xue Teng	Depth perception under scaled motion parallax in virtual reality
		Robert Codd-Downey	Vision based diver-robot interaction at depth
		Sarah Vollmer	Forking paths: The living history of mutable sound

Wednesday, June 16th - continued	**Breakout sessions 11:20am – 1pm**	**VVF Session #9** Chair: George Tomou **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Parisa Abedi Khoozani	Mechanisms for integrating allocentric and egocentric information for goal-directed movements: a neural network approach
		Gifty Asare	Electrophysiological correlates of the effect of the colour red on response inhibition
		Rezaul Karim	Investigation of feedback and lateral inhibition in ConvNets.
		Maria Koshkina	Towards sports video understanding: Player detection, classification, and tracking
		George Tomou	Functional connectivity of transsaccadic perception: Evidence from fMRI and graph theory analysis
	1 – 2pm	**Lunch @Gather Town** **Link: https://gather.town/i/leCDDqJ6; Passcode: 3291** Chat with presenters, hang out with colleagues and friends	
	2 – 4pm	**Workshop #3: Impactful research writing (see p.79) scientific papers and knowledge dissemination** **Dr. Gunnar Blohm, Queen's University** **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	

VISTA – CVR
Virtual Vision Futures

Thursday, June 17th	**Early Session 8:30-9:20am**	**IRTG Session #7** Chair: Frieder Hartmann **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Moritz Schubert	The bayesian causal inference of body ownership model relies on unrealistic parameter choices
		Shanaathanan Modchalingam	Factors affecting implicit motor learning
		Marie Mosebach	Linking signal relevancy and reliability in somatosensory predictions
	Breakout Sessions 9:30 – 11:10am	**IRTG Session #8** Chair: Raphael Gastrock **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Ambika Bansal	The effects of speed and direction of self-motion on visual odometry
		Kevin Hartung	Where are the fish? Advantage from contextual cueing in foraging tasks
		Pierre-Pascal Forster	What can we expect from embodiment and presence ratings: An online experiment
		John Jong-Jin Kim	Can you place your phone in a picture? Determining the distance and size of an object in a 2D representation of a scene
		Ryan Cortez	Frequent cannabis-use has no long-term effects on visuo-motor and visuo-cognitive performance
		VVF Session #10 Chair: TBD **Zoom link: https://tinyurl.com/VirtualVisionFutures2**	
		Aaron Tucker	Solving the conflict between breathability and masked faces within facial recognition technologies
		Diego Fontanive	Implementing metamemetic thinking as an approach to implicit bias reduction
		Marcus Gordon	Live coding volumetric displays
		Rozhan Khalajzadeh	Glaucoma and IOP in swimming goggles
		Noa Yaari	How I think vision when I paint

VISTA – CVR
Virtual Vision Futures

Thursday, June 17th - continued	**Final Session 11:20am – 1pm**	**VVF Session #11** Chair: Keyi Liu **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		Veronica Nacher	Visual response fields in DLPFC during a head-unrestrained reach task.
		Zoha Ahmad	Unilateral cortical resection of both visual pathways alters action but not perception in a pediatric patient with pharmaco-resistant epilepsy
		Keyi Liu	Instance segmentation with sparse distance transform map
		Spencer Ivy	The limits and future of holistic visual processing in visual expertise
		Sanjida Sharmin Mohona	Effects of chromatic aberration compensation on visibility of compression artifacts
	1pm	**Closing remarks: Dr. Doug Crawford** **Zoom link: https://tinyurl.com/VirtualVisionFutures1**	
		End of conference social @Gather Town **Link: https://gather.town/i/leCDDgJ6; Passcode: 3291**	

Presentations

Zoha Ahmad
Session: VVF11 (Thursday, June 17th, 11:15am)

Wavelet flow: Fast training of high-resolution normalizing flows

The cortical human visual system consists of two major pathways, ventral pathway that subserves perception and a dorsal pathway that subserves visuomotor control. These pathways follow dissociable developmental trajectories, and, accordingly, might be differentially susceptible to neurodevelopmental disorders or injuries. Previous studies have found that children with cortical resections of the ventral visual pathway retain largely normal visuoperceptual abilities. Whether visually guided actions, supported by computations carried out by the dorsal pathway, follow a similar pattern remains unknown. To address this question, we examined visuomotor and visuoperceptual behaviors in a pediatric patient, TC, who underwent a cortical resection that included portions of the left ventral and dorsal pathways. We collected data when TC used her right and left hands to grasp blocks that varied in width and length across their width and, separately, to estimate the width of the same blocks perceptually. TC's perceptual estimation performance was comparable to that of controls, independent of the hand used. In contrast, relative to controls, she showed reduced visuomotor sensitivity to object shape which was more evident when she grasped the objects with her contralesional right hand. These results provide evidence for a striking difference in the reorganization profiles of the two visual pathways. This difference supports the notion that two pathways exhibit differential susceptibility to neurodevelopmental disorders.

Tasfia Ahsan
Session: VVF3 (Monday June 14th, 11:15am)

Perceived depth modulates perceptual resolution

Humans constantly use depth information to support perceptual decisions about object size and location in space, as well as planning and executing actions. It was recently reported that perceived depth modulates perceptual performance even when depth information is not relevant to the task, with faster shape discrimination for objects perceived as being close to the observer. However, it is yet to be determined if the observed "close advantage" reflects differences in psychophysical sensitivity or response bias. Moreover, it is unclear whether this advantage is generalizable to other viewing situations and tasks. To address these outstanding issues, we evaluated whether visual resolution is modulated by perceived depth defined by 2D pictorial cues. In a series of experiments, we used the method of constant stimuli to measure the precision of perceptual judgements for stimuli positioned at different perceived distances. In Experiment 1, we found that size discrimination was more precise when the object was perceived to be closer to the observers. Experiments 2a and 2b extended this finding to a visual property orthogonal to depth information, by showing superior orientation

discrimination for “close” objects. Finally, Experiment 3 demonstrated that the close advantage also occurs when performing high-level perceptual tasks such as face perception. Taken together, our results provide novel evidence that the perceived depth of an object, as defined by pictorial cues, modulates the precision of visual processing for close objects

Assel Al-Bayati

Session: VVF2 (Monday, June 14th, 11:15am)

Frequent cannabis-use has no long-term effects on visuo-motor and visuo-cognitive performance

Since the legalization of recreational use of cannabis took effect in Canada, many questions have been brought forward regarding its immediate and sustained effect on daily tasks. To investigate the effect of cannabis on various brain functions, we created a battery of cognitively demanding, visual-spatial and visual-motor tasks. Here, we discuss preliminary findings of four tasks. (1) Serial visual search task which assesses visual attention. Task performance is analyzed by comparing RTs on target present/absent trials for each array set size (6,12,18) in frequent cannabis users (N=45) and non-users (N=172). We found no difference in performance between frequent cannabis users and non-users which indicates that frequent cannabis-use does not impair visual attention. (2) Speeded Go/No-Go task (80% go, 20% no-go), which measures motor impulse inhibition. Accuracy on the task is measured by computing d-prime, as well as RTs on hit/false alarm trials in frequent cannabis users (N=33) and non-users (N=139). Our findings indicate that frequent cannabis users and non-users performed similarly, except that frequent cannabis users were faster to respond. This suggests an absence of a negative effect of cannabis on impulse inhibition. (3) Spatial N-Back task (1, 2, & 3-Back) which assesses working memory/working memory capacity. Performance on this task is analyzed in the same manner as the previous task in frequent cannabis users (N=37) and non-users (N=137). We found that frequent cannabis users were more accurate in 1-Back and they were faster to make errors. This may suggest that frequent cannabis-use is not associated with working memory impairments. (4) Trail-Making test which assesses executive functioning. Performance is analyzed by measuring movement time in frequent cannabis users (N=45) and non-users (N=134). We found that frequent cannabis users were faster than non-users. This finding suggests that frequent cannabis-use does not impair executive functioning. While there might be immediate effects of cannabis-use, our preliminary results show that frequent cannabis-use does not impair visuo-motor and visuo-cognitive functioning.

Katerina Andriopoulos

Session: VVF1 (Monday, June 14th, 9:30am)

Sex differences in visuospatial mental rotation persist under 3D VR conditions

The classic Shepard and Metzler (1971) mental rotation task (MRT) showed a male advantage for visuospatial mental rotation of block structure images. This finding has been replicated numerous times

although one study rendered semi-immersive VR stereoscopic block structure images which eliminated this sex difference (Parsons et al., 2004). In that study, they generated 3D images with the ImmersaDesk stereo-goggle system and found that there was no difference in male and female ability to manually rotate a virtual object to the same spatial orientation as the previously seen target object. Here, we sought to re-examine potential sex differences in mental rotation ability using the original Shepard and Metzler task which required pure mental rotation and did not allow for manual rotation of a 3D image. We developed a novel VR mental rotation task (VRMRT) to re-examine mental rotation using 3D stimuli using more capable, more modern, VR equipment (HTC VIVE). Our hypothesis was that using virtual reality to generate three-dimensional depth information in the block structures for mental rotation would yield sufficient additional structural information for the female observers to better complete the visual mental rotation task and thereby reduce the previously observed male advantage We developed the stimuli in VR using a best-approximation protocol of the Vandenburg and Kuse adaptation of the original Shepard and Metzler task. Twenty-three female and 23 male participants were asked to indicate which two of four spatially rotated images were the same as the target image. They were given unlimited time to complete the entire set of stimuli which consisted of 20 trials. We measured proportion correct as a function of participant sex. Despite the VR set-up, we found a large male advantage, greater than is typically reported. At first blush, these results stand in contrast to Parson's et al (2004) VR study which found no sex differences. However, because that study allowed participants to manually rotate the images it was measuring a visuomotor spatial ability layered on top of mental rotation rather than pure visual mental rotation. Thus, the male advantage in pure mental rotation ability appears to persist even when presented in VR.

Gifty Asare
Session: VVF 9 (Wednesday, June 16th, 11:15am)

Electrophysiological correlates of the effect of the colour red on response inhibition

Blizzard et al. (2017) showed that response inhibition was sped up for red over green without influencing response execution in the stop-signal task. Due to the differential effect between inhibition and execution, they suggested that visual associations for color affected executive functions rather than early visual processing. In this study, we used electroencephalography (EEG) to determine if color does affect the neural circuitry underlying response inhibition in prefrontal cortex. We hypothesized that the frontal N200 component, which has been shown to mediate response inhibition, would be modulated by color automatically. We recorded EEG from 40 young adults while they performed a color-based go/no-go task. The behavioral results showed that participants were more successful at inhibiting their motor responses when the no-go color was red compared to green, yellow, or blue, matching prior studies of response inhibition using the stop-signal task. Consistent with behavior, we found that the amplitude of the frontal N200 was attenuated during successful response inhibition on red no-go trials compared to trials with successful inhibition of green, yellow, or blue no-go stimuli. These findings show

that the color red is preferentially weighted by the prefrontal neural circuitry responsible for response inhibition.

Domenic Au

Session: VVF 6 (Tuesday, June 15th, 11:15am)

The impact of binocular capture on monocular content in augmented reality

Stereopsis is arguably the most precise depth cue available to guide our interactions with both real and virtual three-dimensional (3D) environments. This binocular depth information must be integrated with monocular depth cues such as occlusion, motion, parallax, and perspective (Howard & Rogers, 2012). In the natural environment, there is typically good correspondence between different depth cues. However, the same is not true of virtual environments. In the case of augmented reality (AR), virtual content must be overlaid on real-world surrounding objects. The simultaneous visibility of the physical world in AR headsets means that errors or misalignments between virtual and real-world stimuli will be noticeable and potentially disruptive. In other use cases, monocular imagery is overlaid on the binocular view of the real scene (e.g., heads-up displays). In this type of viewing scenario, the monocular imagery may be affected by nearby physical objects. An extreme example of this is 'Binocular Capture'; a phenomenon where the perceived depth of monocular information is captured by the depth of nearby binocular objects (Erkelens & Van Ee,1997; Shimono et al., 2005; Raghunandan, 2014). If monocular stimuli rendered in AR displays are captured by the depth of surrounding real-world surfaces, it can potentially affect user experience as many AR displays only present augmented stimuli to one eye. Here, we present results of a preliminary study where we evaluate binocular capture of monocular stimuli displayed in AR using a Microsoft Hololens 1™. Stimuli consisted of two virtual monocular letters ('A's) each superimposed on one of two fronto-parallel real-world surfaces, viewed through the HoloLens. To assess the degree to which capture occurs we capitalized on Emmert's Law. That is, that the perceived size of an afterimage changes when located at different distances despite having a constant retinal size (Emmert 1881; Boring 1940). In our case we predict that if capture occurs, the perceived size of a monocular virtual object will vary directly with the distance of a physical surface. Using a method of constant stimuli, the distance of real-world surfaces was varied, and observers indicated which letter appeared larger. Absolute distances ranged from 46.5 to 70.5 cm in 4 cm steps. Our results confirm that the distance of a real-world surface impacts the perceived size of augmented monocular stimuli. Observers perceived the monocular letters as larger with greater real-world distances and as smaller with shorter real-world distances. This work establishes that monocular images presented in AR will appear to be at the depth of the surface they overlay and that we can use the perceived size as a proxy for the degree of capture in future experiments. Understanding when and why monocular content is captured by real-world objects is important for AR devices designed with monocular displays such as the Google Glass™ and the Integrated Helmet and Display Sight System (IHADSS). If user interaction with such monocular content is required, errors in localization will occur and will vary depending on where the user is looking and changes in the real-world environment.

Bianca Baltaretu
Session: VVF 3 (Monday, June 14th, 11:15am)

A cortical network for trans saccadic perception of object features: An fMRI paradigm

Recently, Dunkley et al. (Cortex, 2016) showed modulations in parietal cortex (i.e., supramarginal gyrus, SMG) for transsaccadic changes in object orientation. However, this result did not generalize to other features, such as spatial frequency (Baltaretu et al., Sci. Rep., 2021). We anticipated that the fundamental difference between these results was the detection of transsaccadic changes in object orientation vs. shape. To test this, we used a double-dissociation fMRI task. Participants fixated a small cross 15.4° left or right of centre, where an object was subsequently presented (rectangle, barrel-shaped, or hourglass-shaped), oriented at ±45° from vertical. After this, the fixation cross either remained in the same position (Fixation condition) or shifted to the other side (Saccade condition). Then, either the same object would appear at the orthogonal orientation (Orientation change condition) or one of the other two objects would appear at the same initial orientation (Shape change condition). Button press was used to indicate whether object orientation or shape had changed across the two stimulus presentations in a given trial. Results indicate that cortical modulations were larger for saccades (than fixations) in medial early occipital regions (e.g., cuneus), extending dorsally into parietal cortex (i.e., dorsal precuneus) and ventrally into early-to-intermediate occipital regions (i.e., lingual gyrus, LG). When looking for a feature change (Orientation > Shape) specifically in cortical regions showing saccade-related (Saccade > Fixation) preferences, we found modulations within medial occipital cortex (i.e., cuneus). Functional connectivity analysis using the saccade and feature-sensitive cuneus as a seed region indicated significant communication with early-to-intermediate visual (e.g., LG, superior occipital gyrus), object-relevant (e.g., medial occipitotemporal sulcus, transverse occipital sulcus), and oculo/sensorimotor (i.e., superior parieto-occipital cortex) regions. These results suggest that differences in object features that occur across saccades may also appear at the cortical level (predominantly within medial occipital cortex, extending in both directions toward parietal and temporal cortex)

Ambika Bansal
Session: IRTG 8 (Thursday, June 17th, 9:30am)

The effects of speed and direction of self-Motion on visual odometry

Introduction: Moving through an environment generates relative movement between an observer and their environment known as optic flow. Optic flow alone can generate the sensation of self-motion, known as vection and, in the presence of adequate scale information, can provide a cue for distance traveled although such distance estimates can be imprecise and inaccurate. When moving to the location of a previously seen target, the further the intended target distance, the more people tend to undershoot its location (Redlick et al. 2001; Lappe et al. 2007). This phenomenon has been modelled as resulting from a leaky spatial integrator (Lappe et al. 2007). The model postulates that perceived

travel distance results from integration over distance and is independent of travel speed. However, it is still unclear whether there is an effect of speed. Speed effects would imply integration over time as well as space. In this proposed study, we will measure perceived linear travel distance for a range of speeds and distances. Previous research from this lab investigating visual odometry and the leaky spatial integrator model has so far only examined forward translational movements (Harris et al. 2012; McManus et al. 2017) which raises the question of how well the model predicts travel distance during vertical self-motion or during translational movements backward. We will therefore examine perceived travel distance in these directions also.

Methods: The experiment will be done in virtual reality using visually induced self-motion. Participants will see a target in a horizontal corridor or in a vertical shaft, either above, below, in front, or behind them. When needed, they will turn to view the target, and then return to looking straight ahead. The target will then disappear and motion towards the target's position will be simulated with optic flow. Participants will indicate when they have arrived at the target's location by pressing a button. There will be 4 target distances, 3 speeds, and 4 directions of self-motion: upward, downward, forward, or backward. Data will be analyzed in terms of gain (perceived/actual distance) and fitted with a modified version of the leaky spatial integrator model that will include a speed term.

Expected results: We expect that gains will be higher for vertical and backward motion compared to forward motion, and that a small effect of speed will be found, especially for slower speeds.

Discussion: Our findings will extend the leaky spatial integrator model and further our understanding of how visual motion signals contribute to estimates of travel distance.

References:

1. Harris LR, Herpers R, Jenkin M, et al. (2012) The relative contributions of radial and laminar optic flow to the perception of linear self-motion. J Vis 12:7–7.
2. Lappe M, Jenkin M, Harris LR (2007) Travel distance estimation from visual motion by leaky path integration. Exp Brain Res 180:35–48.
3. McManus M, D'Amour S, Harris LR (2017) Using optic flow in the far peripheral field. J Vis 17:1–11.
4. Redlick F, Jenkin M, Harris L (2001) Humans can use optic flow to estimate distance of travel. Vision Res 41:213–219.

Rebecca Barnstaple

Session: VVF 5 (Tuesday, June 15th, 11:15am)

Visuomotor integration in dance-based learning: Applications for clinical populations

Dance involves the acquisition of motor sequences through observation and repetition, requiring complex visuomotor integration and enhanced attention to spatial/temporal targets on the part of the learner. We are ecologically investigating aspects of dance-based learning to better understand potential applications for neurorehabilitation. Previous studies from members of our group have shown

that dance-based learning over 8-months produced BOLD signal changes in SMA using fMRI (Bar & DeSouza 2016) and resting state alpha power increases in frontal cortex post-dance in people with Parkinson's disease (PwPD) compared to controls (Levkov et al 2014). Our group has also found that PwPD who attended weekly dance classes over 3-yrs show a slowdown in disease progression (DeSouza & Bearss, 2018). To better understand how dance-based learning contributes to rehabilitation in neurodegenerative diseases, our current study uses mobile neuroimaging (MoBI) to assess aspects of motor learning and visuomotor integration while subjects (n=17) learn a 30-sec choreography (Barnstaple et al 2020). The choreography was specifically designed to include elements common to many dance forms without referencing a specific dance style, image, or affect, and synthesized music unfamiliar to all participants was the training stimulus. Recordings were conducted at the Berlin Mobile Brain/Body Imaging lab (BeMoBIL) with motion capture data synced with continuous recording of wireless mobile EEG (Brain Products ActiCap; 128 electrodes) in a dedicated 150 m2 lab space. Movements through space were recorded using 10 HTC Vive trackers running on Steam VR. 30-sec trials included (i) watching VIDEO (4 times), (ii) LIVE performances of the choreography (6 times), (iii) moving with the teacher (LEARN; 3 to 20 times until 80% criteria was reached rated by experimenter, (iv) imagining performing from a first-person perspective (IMAGINE; 6 times), and (v) finally performing in space (PERFORM; 6 times). Music was rated for affective valence pre/post motor learning. All participants reached our target of 80% or higher accuracy in reproducing the movement sequence within 20 LEARN trials, with expertise in dance resulting in fewer trials to reach criteria. Our current analysis focusses on matching behaviour with brain activity to understand how specific changes in neural dynamics may be elicited during the process of watching, learning, and performing dance

Brandon Caie
Session: IRTG 4 (Tuesday, June 15th, 9:30am)

Predictive gating of evidence accumulation

We set expectations about the future by finding patterns in the past; these expectations are combined with evidence when making choices. Previous models assume that evidence is accumulated over a baseline expectation that is fixed within trials, so it is unclear how evidence accumulation would unfold if expectations evolved dynamically over shorter timescales. Participants made eye movements as fast as possible to either of two free choice targets presented at a randomized time from trial onset and from each other. We observed a strategy in which participants balanced reacting to target onset with predicting when it would occur, resulting in anticipatory responses. A sequential analysis revealed that these anticipatory responses were influenced by expectations from previous trials. We hypothesized that this joint dependency may be explained by casting the onset of evidence accumulation as a

predictive process. Predictive gating formalizes this hypothesis within the actor-critic framework, where the onset of evidence accumulation is controlled by both ongoing visual information and a prediction of its timing. Fitting an Ornstein-Uhlenbeck process to data revealed that pre-target gating of the onset of evidence accumulation is necessary to describe anticipatory responses, and the parameters of the gating process must be updated across-trials to account for the joint influence of delay time and choice history. Together, this suggests that the gating of evidence accumulation is set within the context of a prediction for the timing of sensory information

Huiqin Chen
Session: VVF 3 (Monday, June 14th, 11:15am)

Perceptual hysteresis in the categorization of complex scenes

Hysteresis is the dependence of the previous state in a system. In psychophysics, hysteresis refers to the phenomenon that the perception of stimuli is influenced by prior stimuli. Here we ask that when images of scenes are changing continuously, how will participants make categorical judgements and that whether hysteresis will also be found in scene categorization tasks. This study used HiGAN to generate synthesized scene images of different real-world scene categories, and participants were shown sequences of synthesized images that gradually transit between categories. Each transition between two categories is repeated in two opposite directions. Participants were asked to report when they perceived a change in terms of scene categories. Abrupt trials (with sudden shifts between scenes) were set to measure the response time and decision-making time during the tasks. In this experiment, we found that after removing the effect of response time, we still found a significant difference of responses in opposite directions which can be explained by the hysteresis effect. This means that participants tended to stay at the current judgement of scene category until there were significant enough changes happened in the stimuli. Future experiments on a larger scale of scene categories with complicated transitions can provide a deeper understanding of the mechanism, as well as studying the neural activity pattern on brain areas specialized for scene perceptions when resolving perceptual ambiguities

Thomas Chen
Session: VVF 5 (Tuesday, June 15th, 11:15am)

Computer vision for social good: Adapting to the changing climate and extreme weather events

Due to climate change, the frequency and intensity of natural disasters continue to increase. To adapt to this crisis, a number of artificial intelligence approaches are useful. In the field of machine vision, we train deep neural networks (DNNs) on imagery to assess the impact on infrastructure and people. Particularly, there are two sources of imagery data that have emerged to be useful in this field: satellite imagery and social media imagery. Multitemporal satellite imagery aids in the training of change detection-enhanced convolutional neural networks for damage assessment of structures using the earth observation data. Social media, which is an interesting source due to the prevalence of its use during times of crisis for communication among family and the world, provides insights into real-time data from on the ground. In both cases, training artificial neural networks on the data yield results that pertain to identification, severity classification, and semantic segmentation of the damage incurred by devastating natural disaster events, such as wildfires, hurricanes, and floods. In this talk, we discuss this burgeoning area of research, compare social media imagery and satellite imagery as training data for the deep learning-based computer vision models, and assess the deployability of these technologies in the real world to aid in the allocation of resources and personnel in a timely manner during and post-disaster.

Tenzin Chosang
Session: VVF 7 (Wednesday, June 16th, 9:30am)

Evaluating the perceptual efficiency of shape representation models

Our visual system receives an enormous volume of patterns and signals, but we have only a limited number of neurons to encode them. According to the efficient coding hypothesis, the human brain handles this constraint on information processing by taking advantage of statistical redundancies in our visual environment to recode visual data into a more compact form (Kersten, 1987, Simoncelli & Olshausen, 2010). A particular instantiation of this hypothesis predicts that the visual system transforms the perceptual input into ‘sparse’ representations that require only a small number of neurons to encode any particular input. While sparse components of natural image patches have been found to predict the structure of receptive fields in early visual cortex (Olshausen & Field, 1996), these same sparse components have been found to be more statistically interdependent than alternative codes (Bethge et al. 2018). Thus, the exact role of sparse coding in the human visual system remains controversial. This prior work has focused almost exclusively on coding in early visual cortex; little is known about the role of efficient coding in mid-level visual areas coding object shape. My objective is to investigate whether shape representation in the human brain, which plays an important role in object identification (Elder et al. 2018), can also be understood in terms of sparse and/or efficient coding.

Method: I propose to employ a psychophysical method inspired by Shannon's guessing game, first used by Kersten (1987) and then adapted by Bethge et al. (2007) to study the perceptual redundancy in image patches. To adapt the method to shape, I will first use the animal dataset of Bai et al. (2009) and recode them into each of these representations. In the simplest form of the method, one of these shapes will be displayed but with one shape component set to a random gain. The observer's task will be to adjust the gain of this component to its correct value. Better performance on this task will indicate a higher degree of perceptual redundancy between the shape components of the representation, and thus represent evidence against this representation as a model for neural shape representation under the efficient coding hypothesis. In this way I will evaluate a number of shape representation models, including Fourier descriptors (Granlund, 1972), shapelets (Dubinsky & Zhu, 2003), Formlets (Grenander et al. 2007, Oleskiw et al. 2010, Elder et al. 2013) and sparse shape components (Clément & Elder, 2018).

Robert Codd-Downey
Session: VVF 8 (Wednesday, June 16th, 11:15am)

Vision based diver-robot interaction at depth

Current methods for human robot interaction (HRI) in the underwater domain seem antiquated in comparison to their terrestrial counterparts. Fiducial tags and custom-built wired remotes are common solutions in underwater HRI, but such approaches have numerous drawbacks. SCUBA divers make regular use of hand signals to communicate under water. Can robots leverage this prior knowledge to more effectively communicate with divers using a familiar medium. This ongoing research describes a method to perform domain specification recognition using a series of different neural network models to reduce cpu load.

Ryan Cortez
Session: IRTG 8 (Thursday, June 17th, 9:30am)
Frequent cannabis-use has no long-term effects on visuo-motor and visuo-cognitive performance
Since the legalization of recreational use of cannabis took effect in Canada, many questions have been brought forward regarding its immediate and sustained effect on daily tasks. To investigate the effect of cannabis on various brain functions, we created a battery of cognitively demanding, visual-spatial and visual-motor tasks. Here, we discuss preliminary findings of four tasks. (1) Serial visual search task which assesses visual attention. Task performance is analyzed by comparing RTs on target present/absent trials for each array set size (6,12,18) in frequent cannabis users (N=45) and non-users (N=172). We found no difference in performance between frequent cannabis users and non-users

which indicates that frequent cannabis-use does not impair visual attention. (2) Speeded Go/No-Go task (80% go, 20% no-go), which measures motor impulse inhibition. Accuracy on the task is measured by computing d-prime, as well as RTs on hit/false alarm trials in frequent cannabis users (N=33) and non-users (N=139). Our findings indicate that frequent cannabis users and non-users performed similarly, except that frequent cannabis users were faster to respond. This suggests an absence of a negative effect of cannabis on impulse inhibition. (3) Spatial N-Back task (1, 2, & 3-Back) which assesses working memory/working memory capacity. Performance on this task is analyzed in the same manner as the previous task in frequent cannabis users (N=37) and non-users (N=137). We found that frequent cannabis users were more accurate in 1-Back and they were faster to make errors. This may suggest that frequent cannabis-use is not associated with working memory impairments. (4) Trail-Making test which assesses executive functioning. Performance is analyzed by measuring movement time in frequent cannabis users (N=45) and non-users (N=134). We found that frequent cannabis users were faster than non-users. This finding suggests that frequent cannabis-use does not impair executive functioning. While there might be immediate effects of cannabis-use, our preliminary results show that frequent cannabis-use does not impair visuo-motor and visuo-cognitive functioning.

Jonathan Coutinho
Session: IRTG 5 (Wednesday, June 16th, 8:30am)

Luminance evoked pupil response dynamics

Pupil responses are commonly used in clinical assessments and psychological research. Making accurate inferences from pupil measurements requires precise knowledge of the underlying system dynamics, yet this remains poorly quantified. Here we quantify pupil response dynamics in healthy adults (N=10, 5 female, age range: 19-26 years old) during large-field step changes in luminance on a computer monitor, randomly selected between 1 to 43 cd/m2. As commonly reported, we observed a linear relationship between changes in steady-state pupil size and changes log-luminance. Participants varied in their average pupil size (across luminance conditions), and larger average pupil size correlated with increased sensitivity of pupil size change per unit change in log-luminance. When analyzing trial-by-trial pupil dynamics, we observed a saturating nonlinear relationship between peak velocity and diameter change, with dramatic asymmetries between constriction and dilation. Peak dilation velocity was generally small and increased linearly with increasing dilation amplitude (shallow slope). Peak constriction velocities were much larger and initially increased steeply with increasing constriction amplitude before saturating with larger constriction amplitudes. Finally, we explore how these relationships in pupil response dynamics inspire/constrain future models of pupil control mechanisms.

Ben Cuthbert
Session: IRTG 3 (Tuesday, June 15th, 8:30am)

Visual working memory models do not generalize to the whole-report task

The whole-report task is an increasingly popular way to test working memory storage limitations, but it remains unclear whether current working memory models can accurately capture whole-report behaviour. Here, we analyzed two whole-report datasets to test whether predictions and assumptions made by prominent working memory models successfully generalize to the new paradigm. In both datasets we found strong evidence for statistical dependencies between reports of concurrently-presented stimuli, which cannot be explained by models that assumed independent storage. We then turned to working memory models that assume "ensemble" or hierarchical storage and found no evidence for the biases that they predict. Finally, we compared recall for colour and orientation stimuli, and found that although aggregate error distributions were similar, within-trial report behaviour was not. Together these results suggest that working memory models assuming independent storage, ensemble or hierarchical storage, or equivalence between stimulus modalities cannot fully capture whole-report behaviour. We point out that most models of working memory storage were developed in the context of a single stimulus report per trial and suggest that care should be taken when integrating results across different working memory paradigms into a single, unified theory.

Sebastian D'Amario
Session: VVF 2 (Monday, June 14th, 11:15am)

The influence of demographics, lifestyle, and external events on cognitive and motor tasks

Demographics, lifestyle, and experience can affect performance on everyday tasks. This is particularly clear after trauma but may also manifest as individual differences in laboratory tasks. As a baseline for future work in special populations, we created a battery of online visuo-cognitive and visuo-motor tasks and collected data from over 100 undergraduate participants. Here we present preliminary findings on 3 of the tasks in the battery. We have measured performance on (1) a Go/No-Go task (20% no-go trials), to assess inhibition or impulse control using sensitivity (d') and reaction times. We found no effects on d' but hit RT increases somewhat with age and is lower in video-game players (N=35) than non-players (N=63). In (2) a reach task with mirror-reversed feedback, we used path length and movement time to assess cognitive-motor integration. We found shorter movement times in males (N=27) compared to females (N=78), and in video-game players (N=25) compared to non-players (N=59). We also found shorter path lengths in people with a concussion history (N=8) compared to those without a concussion history (N=97). Finally, we assessed selective attention (3) using reaction times in a serial visual search task with 3 set sizes (6, 12, or 18 items), and in trials where the target was either present or absent. We found no difference in selective attention between males and females

or between video-game players and non-players. Our findings confirm that demographics and external events contribute to individual differences in task performance. We can now start to test performance on these tasks in special populations (e.g. in healthy ageing) and understand how much individual differences within these populations contribute to performance and how this affects everyday life.

Diego Fontanive
Session: VVF 10 (Thursday, June 17th, 11:15am)

Implementing metamemetic Thinking as an approach to implicit bias reduction

By approaching the problem of implicit bias through a memetic viewpoint we have to consider how our cognitive information processing is conditioned by countless of memetic thought processes which replicate themselves acritically. "Memetic" refers to units of culture for instance, as proposed by researchers such as evolutionary biologist R. Dawkins, D. Dennett, Dr. S.
Blackmore, etc. Dawkins rightly called memes "viruses of the mind' To analyze memes critically and constructively what's required is a counter-intuitive approach focused on differentiating the very activity of thinking from the mere production of thoughts. The premise is that thought and thinking are two different things. An important inquiry regarding this approach focuses on whether is possible or not to think about our thoughts without using the same thoughts that have created those thoughts. We will call this approach "metamemetic thinking" which refers to the development of thinking skills capable decoding critically the fallaciousness of thought patterns before adopting them. The modern human world we live in offers a spectrum of available distractions and entertainments as we never had in the past. Another element is represented by the fact that while we always had a biological self (genes) and a psychological self (thoughts), today we also have a digital self, and our sense of identity and psychological security is getting more focused on feeding the digitized self. Many indicators seem to show how we are moving towards a society focused on entertainment, polarized viewpoints and general superficial thinking and this highlights the urgency of implementing our thinking skills. Critical thinking for instance, which does not seem to work, should be combined with the understanding of memetics and the application of high order multilogical and metamemetic thinking skills especially considering how AI development is and will be more and more interfaced with the human psycho-social experience.

Pierre-Pascal Forster
Session: IRTG 8 (Thursday, June 17th, 9:30am)

What can we expect from embodiment and presence ratings: An online experiment

The relation to our body and to our environment is captured by the terms embodiment and presence, respectively. Both play an important role in virtual reality (VR) applications, spanning clinical treatment, training purposes, and gaming. Despite the relevance of both concepts, it is so far unclear how they relate to each other. Recent findings also challenge embodiment research by suggesting participants might know the hypotheses of experiments and respond accordingly. Here, we present data of an online experiment which aimed to control for possible confounds in a later laboratory VR experiment concerned with the relation between embodiment and presence. In this online experiment, material was presented to participants which showed and explained the experimental setup of the laboratory experiment. The same questionnaire items as in the laboratory experiment were used for the online experiment. Participants were asked to respond according to what they think they would answer, if they participated in the laboratory experiment. Our results show that participants' responses in the online experiment were similar to what one would expect in the laboratory experiment. This suggests that results in laboratory settings might be influenced by participants' expectations. As it is important to assess the strength of this effect, in a second step, we plan to compare results from the online experiment to the same experiment conducted in the laboratory.

Elena Fuehrer
Session: IRTG 4 (Tuesday, June 15th, 9:30am)

Sensorimotor predictions lead to sensation-specific tactile suppression

Tactile sensations on a moving body part that occur while a movement is being planned or executed are perceived as less intense than the same sensations occurring at rest. This phenomenon has been explained by predictive mechanisms. Based on efference copy signals, internal forward models are assumed to predict the sensory consequences of a movement, and the predicted sensory consequences are suppressed. Previous research shows that both the predictability of movement-relevant object features, and the movement-relevancy of somatosensory information modulate such suppression of tactile sensations. We examined whether tactile suppression is specific to the nature of the predicted sensory consequences, or whether it results from a general cancellation process. Participants were instructed to move the tips of their index fingers across textured objects at a designated speed. Depending on the spatial period of the texture, these movements caused participants to experience either a low (40 Hz) or a high (240 Hz) frequency vibration at their fingertips. Objects were presented in a blocked manner, so the sensory movement consequence i.e., the vibration frequency, was predictable (either low or high, depending on the block). To quantify tactile suppression, participants had to detect additional vibrotactile probes of varying intensities that were applied at the base of the moving index finger around movement onset. These vibrotactile probes also had either a

low (40 Hz) or a high (240 Hz) frequency, matching (congruent) or mismatching (incongruent) the predicted sensory consequences of the movements. Compared to measurements taken on the same participants at rest, and in line with previous work, detection thresholds of the vibrotactile probes were elevated during movement, clearly indicating tactile suppression. Meanwhile, detection precision was not hampered during movement. Interestingly, congruent probes matching the predicted movement consequence received greater suppression overall than incongruent probes, but this was systematic for the low frequency probes only, with a clear effect and a trend for the low and high frequency probes respectively. These results show a modulation of how external vibrotactile probes are suppressed depending on the predicted movement consequence. Altogether, our findings indicate that sensorimotor predictions can lead to sensation-specific tactile suppression and argue against a mere general cancellation process.

Raphael Gastrock
Session: IRTG 1 (Monday, June 14th, 8:30am)

Skill acquisition versus adaptation: Using a mirror reversal task to investigate de novo learning and distinguish it from motor adaptation

People move within constantly changing environments. Such changes may lead to movement errors, which we must process to produce the correct movement. This error processing occurs regardless of whether people are acquiring new motor skills (de novo learning) or adapting well-known movements (motor adaptation). However, these two types of motor learning should develop differently. Furthermore, while reach aftereffects or the persistent deviation in reaches after perturbation removal are typically observed following adaptation, switching between response mappings in de novo learning should not lead to aftereffects. Here, we conducted two experiments that differentiate de novo learning from adaptation and explore the mechanisms of de novo learning further. In experiment 1 (N = 16), we distinguished the two motor learning types by having participants reach to targets while training with two perturbations: a 30-degree visuomotor rotation and a reversal of cursor feedback in the opposite direction of a mirror axis. We matched the movements required to reach the targets in both tasks and found no order effects, suggesting that learning for one perturbation does not affect the other. Although participants countered for both perturbations by the end of only 90 training trials, learning for the rotation task was more gradual while variability in learning was greater for the mirror task. Participants generally took longer to initiate and execute reaching movements in the mirror task. Moreover, participants only exhibited reach aftereffects after completing the rotation task. In experiment 2, participants completed an online version of the mirror reversal task across two sessions. During the first session, we compared whether providing participants with explicit instructions about the nature of the mirror reversal (N = 105) or not providing such instructions (N = 645) affected learning. Two targets were located in the upper-right quadrant of the workspace (30 and 60 degrees in polar coordinates), with the mirror located along the vertical midline axis. Surprisingly, we found that learning occurred quickly, even for the non-instructed participants. Moreover, asymptotic learning of participants in both groups differed depending on target location, and reach aftereffects were not observed. We then had Non-instructed participants from session 1 return for a second session (N = 434; days apart: M = 16.22, SD = 16.41). To assess retention of learning, we first had participants reach with the same conditions as in session 1. We then tested for generalization across the workspace by switching the target locations to either the lower-right quadrant (300 and 330 degrees) or the upper-left quadrant (120 and 150 degrees) of the workspace, with the mirror in the same location. Finally, we had participants switch to their opposite and untrained hand to reach to targets in the upper-right quadrant. We found that participants retained learning from session 1, as they were immediately reaching towards the correct direction relative to the mirror. We also observed that learning generalized to the different target locations and to the opposite hand. These results not only show the behavioral mechanisms underlying de novo learning, but also distinguish it from adaptation.

Caroline Giuricich
Session: VVF 2 (Monday, June 14th, 11:15am)

Target-distractor competition modulates saccade trajectories

Recent studies showed that the similarity between nearby target and distractors affects the curvature in saccade trajectories. In this study, we varied the distance between, as well as the similarity of complex target and distractor objects in a delayed match to sample task to examine their effects on saccade trajectories. At short saccadic reaction times, there was little effect of similarity or distance. At longer SRTs, there was sufficient time for competition between the objects to develop. Saccade curvature was modulated by the target-distractor distance, exhibiting the effects of a spatial suppressive surround as is found in visual processing areas. As the target-distractor similarity decreased, the initial saccade angle away from the distractor increased, reflecting stronger distractor inhibition. Taken together, these results support a stronger role for visual processing driving saccade planning rather than a winner-take-all competition. Target-distractor interactions can be intrinsically measured from the saccade trajectory to the target.

Nedim Goektepe
Session: IRTG 6 (Wednesday, June 16th, 9:30am)

Understanding the temporal dynamics and informational content of foveal feedback in peripheral vision

In an fMRI study, Williams et al. (2008) found that during peripheral discrimination, object information is fed back to the fovea retinotopic cortex. Subsequent behavioral studies reported that peripheral object discrimination is disrupted by an asynchronously presented foveal noise (Fan et al., 2016; Weldon et al., 2016; Yu & Shim, 2016). However, the effective time window of the foveal feedback varies across studies. One possible explanation could be the difference between task demands (Fan et al., 2019). Another reason could be the use of foveal noise with different statistical properties. In the current project, we will explore how noise with different spatial frequencies modulates the temporal dynamics and the efficiency of the foveal-feedback. We are particularly interested in spatial frequencies because the visual system processes information in a coarse-to-fine fashion. Traditionally, overlapping noise masks are most effective when their spatial frequency is matched to the target's spatial frequency (Stromeyer & Julesz, 1972). We hypothesize that the effective time window of foveal feedback is modulated by the processing speed for different spatial frequencies and that noise with a spatial frequency matched to the target's spatial frequency is most effective in disrupting object discrimination.

Marcus Gordon
Session: VVF 10 (Thursday, June 17th, 9:30am)

Live coding volumetric displays

My research at the nd :: StudioLab embodies an exploration into algorithmic composition processes, but also in the making of instruments for both expression and analysis. In direct relation to my dissertation research, this exploration begets a narrative around the epistemological nature of live coding and how I intend to apply it to academic research of systems that further the understanding of human association to nature, more specifically atmosphere. This project demonstrates my thinking behind the use of live coding as an approach to composition and ArtScience research. I present my findings in live coding of music and visuals as a form of archimusic, a conflation of music and architecture. Wanting to explore the idea of live coding with volumetric displays as an approach to archimusic, I create a live coding interface prototype to transform a volumetric display into an instrument for audiovisual performance. The volumetric display used in this research, the IceCube Display, is part of a joint research collaboration between the Wisconsin IceCube Particle Astrophysics Center (WIPAC) at the University of Wisconsin-River Falls and affiliated faculty, students and researchers of nd :: StudioLab at York University. This collaboration seeks to build volumetric scale models of the IceCube Neutrino Observatory at the south pole, and to build a display that can be used as a model for education and outreach.In spirit of this wider objective, the making of an interface that transmits instructions to the display, introduces a collaborative play element to experiencing the IceCube Displays. By demoing and presenting the makeup of this prototype (called Sigview), I highlight various elements of the prototype that makeup the system and their relationships. In addition, an electroacoustic composition perspective is presented that situates the work as visual music and a modular synthesizer of light and sound. **Keywords:** live coding, architectonic media intervention, visual music, audiovisual performance, volumetric displays, algorithmic composition, electroacoustics.

Mylann Guevara
Session: VVF 5 (Tuesday, June 15th, 11:15am)

The neural underpinnings of mental attention capacity in healthy young adults

Working memory (WM) enables the online maintenance and manipulation of information to solve problems and is constrained by a limited mental attention capacity (M-capacity). The Theory of Constructive Operators (TCO) describes M-capacity as a domain-general operator that can boost task-relevant information in the brain, while task-irrelevant information is inhibited, to facilitate successful task performance. The activation and interactions of functional brain networks that underlie M-capacity are not well known. Further, there is conflicting evidence regarding how neural activation is modulated as a function of cognitive load and the domain of information held in working memory. The current investigation used functional magnetic resonance imaging to quantify brain activation in healthy young adults while they performed three different WMtasks designed to test specific hypotheses based on the

TCO. Brain activity was dynamically modulated by cognitive load, regardless of stimulus domain (colors vs. numbers), such that brain regions associated with cognitive control and attention networks showed monotonically increasing activation, while regions associated with the default network showed decreasing activation, as WM load increased, within the range of M-capacity. However, both sets of brain regions showed a complete reversal of this activation pattern at the point where M-capacity was exceeded. Moreover, this relationship interacted with stimulus-domain in other brain regions, with differential activation within the fusiform gyrus(bilaterally) and right middle occipital gyrus depending on stimulus domain. These observations support the TOC and inform a number of future investigations aiming to understand the development of M-capacity in childhood/adolescence and across the adult lifespan. **Keywords:** fMRI; mental attention capacity; working memory; neural networks; default network; cognitive control network; attention network; cognitive load; theory of constructive operators

Sidharth Gupta
Session: VVF 4 (Tuesday, June 15th, 9:30am)

Implementing and integrating contour completion using perceptual grouping

Perceptual grouping is the process of grouping visual elements that have a similar or a unifying pattern into larger, meaningful units. Principles behind perceptual grouping have been extensively studied in human vision and have greatly influenced computer vision algorithms in the past decades. However, significant advances in deep neural networks have made algorithms in perceptual grouping lose favor. Here we revisit and modernize the idea of stochastic completion fields (SCFs), a method for grouping contour elements and completing the illusory gaps between them. We enhance the original SCF algorithm in many ways: making SCF's more efficient to compute, enabling them to not require assignment of sources and sinks, and also building an intuitive data structure that makes completing illusory contours easy. We test our SCF framework on line-drawing datasets and show that it mimics results in human perception. In addition, we demonstrate the framework’s potential in a variety of computer vision applications. We use our SCF algorithm to improve state-of-the-art generative inpainting and contour detection in extremely noisy input models. In a series of benchmarking tasks, we find that integrating SCF in these models helps them achieve statistically significant better SSIM scores than before. Overall, our revisit of stochastic completion fields shows the prominent impact that perceptual grouping can make in helping modern computer vision tasks.

Mahboubeh Habibi
Session: IRTG 4 (Tuesday, June 15th, 9:30am)

A Free viewing paradigm to identify eye movement disturbances in alpha-synucleinopathies versus tauopathy

Clinical assessment of oculomotor (dys-)function is essential in the differential diagnosis of neurodegenerative disorders. Eye movements deficiencies have been studied extensively in Parkinson's disease. Here, we employ video-based eye tracking with instruction-free viewing of video clips to assess oculomotor, pupil, and blink behavior in alpha-synucleinopathies (aSYN) and a tauopathy (TAU).

Methods: We compared patients suffering from aSYN (manifest Parkinson Disease (PD); manifest Multiple System Atrophy (MSA); REM sleep behavior disorder (RBD) - a prodromal stage of aSYN) with patients suffering from the TAU progressive supranuclear palsy (PSP) and healthy age-matched controls (CTRL). Participants sat in front of a video-based monocular eye tracker (Eyelink-1000 Plus, SR Research) in a light-controlled, quiet room. A monitor mounted camera with a 500 Hz sampling rate used to measure eye movements and pupil size. All participants viewed 10 short videos (~1 minute), consisting of 16-17 clippets of 3-5s duration without further instructions. Gaze pattern, saccade, pupil, and blink analysis were segmented by clip boundaries.

Results: Using the fixation pattern of the subjects, we found that PSP had more substantial center bias than other patients. Consistent with the literature, had severe vertical saccade deficits. Horizontal saccade amplitude was also significantly reduced in PSP but they exhibited exaggerated horizontal microsaccades. PSP also showed slower saccade peak velocity. All groups except RBD had reduced saccade frequency compared to CTRL.PD and PSP had longer fixation durations. In addition, PSP blinked less often than CTRL and other patient groups. We examined various eye movement parameters time-locked to onset of the clip change, which induced a strong visual perturbation into the visual system. We observed a number of oculomotor, pupil, and blink behaviors that differed between groups. For saccades, there was a momentary suppression immediately following the clip change followed by a rebound. This rebound in saccade rate was reduced in all groups, especially in PSP, relative to controls. In contrast, microsaccade suppression was reduced in the patient groups relative to controls. Clippet transition could also produce a large change in screen luminance from dim to bright or vise versa, requiring pupil constriction or dilation, respectively. PSP had a significantly smaller constriction than CTRL, but PD and MSA had greater constriction than CTRL. When the clip change produced a decrease in luminance, the resultant pupil dilation was reduced in PSP, but increased in the other groups relative to CTRL. The Synucleinopathy groups (PD, MSA) showed more pupil constriction after luminance increase, whereas the tauopathy group (PSP) showed less constriction. The opposite occurred for luminance increases. The PSP group was also different in vertical gaze palsy and a great center bias. Our results show that changes in saccade and pupil responses differed between Synucleinopathy and tauopathy patients and may be very important for differential diagnosis in early disease stages.

Brittney Hartle
Session: VVF 6 (Tuesday, June 15th, 11:15am

Scaling stereoscopic depth through reaching

Depth estimation from stereopsisis biased under many viewing scenarios and for a range of estimation methods, particularly for virtual stimuli. These distortions are often attributed to misestimates of viewing distance that result in incorrect scaling of binocular disparity. The majority of research on depth scaling has considered only visual cues to distance. However, we do not just look at the world, we interact with objects and in this way may have access to proprioceptive cues to distance. There is evidence that stereopsis aids actions such as prehension; is the reverse also true? We assessed the impact of proprioceptive distance from arm's reach on stereopsis using a ring game that is contingent on accurate absolute distance perception. Observers used hand controllers and their index finger to move rings onto a peg in a virtual environment. They completed the task as quickly as possible while avoiding touching the rings to the peg (errors were signaled via controller vibration). After each block of 5 trials observers were given feedback regarding their completion time and accuracy. To evaluate the impact of this proprioceptive experience we assessed depth magnitude estimation before and after completion of the ring task. Observers were asked to estimate the depth between a rectangle and a reference frame located at the same distance as the peg in an otherwise blank field. We found that depth estimation accuracy and scaling improved with experience. Importantly, in a follow-up experiment we found that this improvement was contingent on performing the reach. Consistent with the assumption that observers underestimate absolute distance, we found that most ring-placement errors were due to underreaches. We conclude that the improvement in depth estimation seen here reflects a cross-modal calibration of visual space that is under appreciated, but potentially important for everyday interactions.

Frieder Tom Hartmann
Session: IRTG 4 (Tuesday, June 15th, 9:30am)

Hand configurations and contact regions of natural human grasps

When we grasp an object, we must not only configure our hand pose appropriately to the size, shape and material of the object, but also select appropriate contact regions on the object surface that allow comfortable and stable grasps. Here, we aim at modelling both grasp configurations and contact regions when participants freely grasp a wide range of natural 3D objects. On each trial, participants were presented one of 31 objects and were asked to pick it up. Using a Qualisys motion capture system, we tracked 21 markers attached to the participants' hand joints, and in doing so acquired high fidelity hand poses across the full trajectory of action execution. We then constructed a grasp space of natural grasp configurations by applying PCA to joint Euler angles attained from the marker data. We also took

advantage of recent advances in 3D hand mesh reconstruction and projected reconstructed hand meshes onto the surface of grasped objects, allowing us to estimate contact regions between the hand and the object. The acquired dataset enables new insights into how we perform natural grasps to interact with objects in our environment.

Kevin Hartung
Session: IRTG 8 (Thursday, June 17th, 9:30am)

Where are the fish? Advantage from contextual cueing in foraging tasks

Patterns and regularities in our environment could give us advantages in pending tasks. Studies on contextual cueing show that task-irrelevant context elements can improve the detection of a relevant target if they are repeatedly presented. In most contextual cueing studies, the location of L-shaped distractors functioned as cue for the position of a single T-shaped target. The repeated distractor locations guide attention to the associated target position, resulting in better performance compared to novel trials, even if the manipulation of the arrangement is unperceived. The present study transfers this concept to a foraging task, where multiple instances of a target could appear, and participants were asked to collect as many as possible of them in an efficient way. The target was a moving fish-like object, the distractors three different kind of fishes. The fishes were swimming around stationary, coral-like objects in differently colored background displays. We assumed that repeatedly presented constellations of corals, background and fishes would result in a contextual cueing effect, so that the participants collect the targets faster in repeated compared to novel patches. In our first experiment a complex combination of background color and coral types served as context in each patch. We assumed that a fixed constellation of four different coral types and the background color could be an adequate context, but first analysis showed no evidence for an advantage for the repeated trials. We conclude that contextual learning requires a more distinct connection between the cueing and the target objects and will test this assumption in further experiments.

Kieran Hussey
Session: IRTG 6 (Wednesday, June 16th, 9:30am)

Familiar size reliably affects size and distance perception in high-resolution virtual reality

The visual system uses a combination of external and internal cues when processing visual scenes. Previous research from our lab has found the familiar size of objects affects size and distance perception, both in reality (when oculomotor cues are absent) and in virtual reality (VR; regardless of whether oculomotor cues are present or absent). However, these studies tested only two objects, so it is possible that familiar size effects could be confounded by the properties of those specific objects.

Here, we examined how familiar size affects size and distance perception in VR by presenting a more diverse range of objects (virtual sports balls) at familiar and unfamiliar sizes under monocular and binocular viewing conditions. Specifically, large sports balls (basketball, soccer ball, and volleyball) were displayed at the average size of small sports balls (baseball, pool ball, and golf ball), and vice versa, in addition to being shown at their familiar sizes. We predicted that physical size and distance perception would be biased towards an object's familiar size when presented at an unfamiliar size. After testing 22 participants (11 male and 11 female), the results were found to support this prediction. For example, when a basketball was made the size of a small sports ball, it was perceived as larger and farther away than it was displayed. Furthermore, the specific visual properties of the objects displayed was shown to have a negligible effect on size and distance perception using a nested statistical analysis, further supporting the conclusion that familiar size influences size and distance perception. Additionally, we replicated a previous finding that size and distance perception under binocular conditions was similar to that during monocular conditions in VR (unlike real-world stimuli where binocular viewing leads to more accurate perception). Because we optimized oculomotor cues to a greater degree than earlier studies using a superior VR headset, this strengthens the interpretation that the lack of a binocular advantage in VR is due to the vergence-accommodation conflict.

Spencer Ivy
Session: VVF 11 (Thursday, June 17th, 11:15am)

The limits and future of holistic visual processing in visual expertise

Studies in the psychology of visual expertise have tended to focus on a limited set of expert domains, such as radiology and athletics. Conclusions drawn from these data indicate that experts use parafoveal vision to process images holistically. In this study, we examined a novel, as-of-yet-unstudied class of visual experts—architects—expecting similar results. However, the results indicate that architects, though visual experts, may not employ the holistic processing strategy observed in their previously studied counterparts. Participants (n = 48, 24 architects, 24 naïve) were asked to find targets in chest radiographs and perspective images. All images were presented in both gaze-contingent and normal viewing conditions. Consistent with a holistic processing model, we expected two results: (1) architects would display a greater difference in saccadic amplitude between the gaze-contingent and normal conditions, and (2) architects would spend less time per search than an undergraduate control group. We found that the architects were more accurate in the perspectival task, but they took more time and displayed a lower difference in saccadic amplitude than the controls. Our research indicates a disjunctive conclusion. Either architects are simply different kinds of visual experts than those previously studied, or we have generated a task that employs visual expertise without holistic processing. The data suggest a healthy skepticism for across-the-board inferences collected from a single domain of expertise to the nature of visual expertise generally. I shall focus on this skepticism,

using it as inspiration for three interpretive theses concerning the place of holistic visual processing in the development of skill and expertise. First, it may be that our conception of the behaviors associated with holistic visual processing is too narrow. Different classes of visual experts holistically process, but they do so by exhibiting distinctive behaviors unique to their respective domains. Second, it might simply be that the holistic visual processing model of visual expertise is limited by domain, and so is not a perfect indicator of whether or not someone is a visual expert. Finally, it may be the case that holistic visual processing is not simply a physical phenomenon but is also very much a mental phenomenon as well. I shall argue that this third interpretation is most likely given our study and other research in its near orbit on holistic visual processing.

Björn Jörges
Session: VVF 6 (Tuesday, June 15th, 11:15am)

Visually simulated self-motion biases the perception of object motion and makes it less precise

When a moving observer views a moving target, the same retinal speeds can correspond to vastly different physical target speeds. For example, when observer and target move in the same direction, the retinal speed of the object is partially cancelled out, and vice-versa. Observers must thus obtain an accurate estimate of their own velocity and subtract it from or add it to the retinal speed elicited by the target to recover the object velocity accurately. Estimating self-motion speed should be facilitated when visual and vestibular cues are congruent and can be integrated without multisensory conflict [1]. When a static observer undergoes only visual motion, compensation is likely to be incomplete, leading to biases in judgments of object speed (Hypothesis 1). Furthermore, self-motion information is noisier than retinal information concerning object motion [2], especially when only visual cues are available [3]. Subtracting noisy self-motion information from retinal motion to estimate target velocity should thus decrease precision (Hypothesis 2). To test these hypotheses, we presented two motion intervals in a 3D virtual environment and asked participants which motion was faster; one in which a target moved linearly to the left or to the right in the fronto-parallel plane, and one that consisted of a cloud of ¬smaller targets travelling in the same direction. The single target moved at one of two constant speeds (6.6 or 8m/s, 6m from the observer), while the speed of the cloud was determined by a PEST staircase. While observing the single moving target, participants were moved visually either in the same direction (congruent), in the opposite direction (incongruent), or remained static. In two control conditions, we tested for possible effects of induced motion: a blank backdrop condition that should elicit self-motion, but no induced motion, and a condition with a moving backdrop and a static observer that should elicit induced motion, but no self-motion. Regarding Hypothesis 1, participants judged target motion during incongruent self-motion as faster than in the static condition, but still compensated for about 85% of self-motion. There was no difference in accuracy between congruent self-motion and the static condition, suggesting a near complete compensation. Regarding Hypothesis 2, we found similarly that

only incongruent self-motion decreased precision, while precision was similar for congruent self-motion and the static condition. In the blank wall condition (no induced motion, only self-motion expected), we found similar results as in the main experimental condition, while the moving wall condition (no self-motion, only induced motion expected) elicited a small bias indicating that there was indeed some induced motion. As expected, the effect of induced motion was opposite to the effect of visually simulated self-motion, that is, compensation for self-motion was likely less complete than the results of our main experimental condition suggested. **Acknowledgements:** LRH is supported by an NSERC discovery grant. BJ is supported by the Canadian Space Agency.

References:

1. Harris et al. (2000). Experimental Brain Research, 135(1)
2. Dokka et al. (2015). Cerebral Cortex, 25(3)
3. Fetsch et al. (2010). European Journal of Neuroscience, 31(10)

Rezaul Karim

Session: VVF 9 (Wednesday, June 16th, 11:15am)

Investigation of Feedback and Lateral Inhibition in ConvNets

My work studies lateral inhibition and top-down feedback in convolutional networks (ConvNets). As a particular task to ground these studies, I currently focus on video object segmentation. Extensive qualitative and quantitative empirical evaluation on standard datasets show the promise of the approach. Based on this success, we continue to work toward state-of-the-art segmentation performance as well extensions to other video understanding tasks, e.g., action detection. Lateral inhibition and top-down feedback are prominent operations in biological vision systems, yet are far less explored in computer vision. In particular, the currently dominant use of ConvNets for computer vision largely has neglected consideration of these operations. This lack is particularly striking, because ConvNets themselves were originally inspired by biological considerations. My work addresses this shortcoming by studying how lateral inhibition and top-down feedback can be incorporated into ConvNets. As a particular task for incorporation of lateral inhibition and top-down feedback, I consider unsupervised video object segmentation. Video object segmentation is concerned with the delineation of salient objects in a video. It has potential to benefit from lateral inhibition in helping to define sharp object boundaries and from top-down feedback in helping to focus processing in salient areas. Moreover, previous work on single image segmentation has shown benefits of feedback [1]. As a backbone ConvNet architecture for my studies, I use SOE-Net [2]. SOE-Net is an analytically defined ConvNet, which has yielded state-of-the-art performance on dynamic texture recognition. Since it has shown strong performance on dynamic texture recognition, it provides a reasonable backbone for delineating different dynamically defined regions for video segmentation. Lateral inhibition is implemented as a multiscale pyramid of difference of Gaussian filtered features. The difference of

Gaussians at each scale helps to contrast feature responses corresponding to the objects at that scale compared to the surroundings. The multiple scales capture features for objects at different scales, which is an important challenge in real-world video data. The top-down feedback mechanism is motivated by the feedback phenomenon found in the mammalian visual cortex. The features from the deeper layers of SOE-Net are fed back to the input of early layers as a guiding signal, i.e., where to focus. The final feature map is then extracted in an iterative manner modulating the input using the feedback. This iterative algorithm has been found to extract more spatially discriminative features at the location of the object of interest. Extensive experiments have been performed on a benchmark video segmentation dataset, DAVIS-2016 [3]. The proposed approaches improve over the vanilla SOE-Net features and generate results competitive with the state-of-the-art. Based on these results we continue to refine our approach, e.g., through incorporating machine learning.

References:

1. Karim, R., Islam, M.A. & Bruce, N.D.B., Recurrent iterative gating networks for semantic segmentation. Proc. Workshop on Applications of Computer Vision (WACV),2020.
2. Hadji, I. & Wildes, R.P., A spatiotemporal oriented energy network for dynamic texture recognition, Proc. International Conference on Computer Vision (ICCV),2017.
3. https://davischallenge.org/

Johannes Keck

Session: IRTG 4 (Tuesday, June 15th, 9:30am)

Spatiotemporal features of emotional body language in social interactions

How are emotions perceived through body language in human interactions? Although emotion perception is a complex phenomenon integrating various features, humans are highly adept at recognizing and expressing emotions through body movements within social interactions. This study used point-light displays of social interactions portraying emotional scenes to address two issues. First, we examined quantitative intrapersonal kinematic and postural body configurations. Second, we calculated interaction-specific parameters and analysed how far both contribute to emotion and valence perception. By using ANOVA and classification trees we investigated emotion-specific differences in the calculated parameters. Further, we applied representational similarity analyses to determine how perceptual ratings relate to intra- and interpersonal features. Results showed that within an interaction, intrapersonal kinematic cues corresponded to emotional ratings, whereas postural cues reflected valence ratings. Furthermore, emotion perception of emotional content was driven by interpersonal orientation, proxemics, the time spent in the personal space of the counterpart, and the balance in motion energy between interacting people. Both the latter and balance in orientation related to valence ratings. Hence, intrapersonal and interpersonal features of emotional body language relate to not only the emotional content in human interactions but also their perception.

VISTA – CVR
Virtual Vision Futures

Rebecca Keller
Session: VVF 4 (Tuesday, June 15th, 9:30am)

(Endogenous) perceptual states are conceptual

It is standard to take the conceptual-nonconceptual distinction to be about the contents of various mental states. However, several authors (e.g., Byrne; Heck) have pointed out that arguments for content non-conceptualism about perception really at best support only a claim about perceptual states, not perceptual content. The state claim is just that conceptual states are those types of states that, to be tokened, require the subject to possess the specifying concepts; nonconceptual states are those types of states that do not require concept possession to token. That is, the distinction is between states that are concept-dependent and those that are concept-independent. Still, though, the distinction between non/conceptual states is often taken to fit on to the perception-cognition distinction; perception is taken to be nonconceptual and cognition to be conceptual. In this project I argue, contra the received view, that at least some types of perceptual states are concept-dependent. Specifically, I argue that any endogenously produced perceptual state—endogenously-cued expectational states and states of feature-based attention are paradigms—will be concept-dependent. This requires first and foremost that I establish that these states are perceptual, which I argue for on the basis of neuroscientific and psychological work on 'perceptual templates' and 'attentional templates'. The claim that an endogenous state requires concepts to token is more straightforward. Consider a typical expectation-cueing paradigm, where a high tone might alert you to expect a red square, for instance. It is hard to see how one could token any anticipatory state in response to the cue if one did not have the requisite concepts: how else would one know what to expect? If my defense of the state view succeeds, there are a number of consequences for our understanding both of perception and of belief. I focus here on the former. For one, we are in a position to reply to a notably strong objection to the state view (raised by Bermúdez, among others): the only reason to think that there is a difference between concept in/dependent states is precisely because there is a difference in content, and so the state view collapses on to the content view. Instead, concept-dependent and concept-independent perceptual states demonstrably do have the same type of content. The difference is in functional profile: where stimulus-dependent states function to represent the present; anticipatory states function to prepare us for the future. Both Byrne and Heck note that the state view is consistent with there being no difference in the sorts of contents between concept-dependent and concept-independent states; here, I demonstrate that this is the case. Conceptuality, then, has to do with the function of a state, not with its contents. A second consequence is that we should view perception as a heterogeneous kind, containing state types that are both concept-dependent and concept-independent. That is, we ought to have a more fine-grained classification of types of perceptual states. Considering the differential concept-dependency of various perceptual states is informative in the overarching project of giving a more fleshed-out taxonomy of perception.

Rozhan Khalajzadeh
Session: VVF 10 (Thursday, June 17th, 9:30am)

Glaucoma and IOP in Swimming Goggles

Although individuals diagnosed with Glaucoma are advised against taking part in many physical activities like lifting heavy weights or exercises that require the inferior placement of the head to the heart, aerobic exercise is proven to lower eye pressure and benefit those suffering from Glaucoma. The aerobic exercise of swimming is often misunderstood and blamed for elevated IOP in individuals with Glaucoma; the IOP is possibly increased through decreased blood flow in the periocular area due to use of swimming goggles and their design. Research shows that "The average IOP before, during and after wearing the swimming goggles were 11.88 ± 2.82 mmHg, 14.20 ± 2.81mmHg and 11.78 ± 2.89 mmHg, respectively. The IOP increased immediately after putting on the goggles ($p < 0.05$) and then returned to normal values immediately after removal ($p > 0.05$)". In addition, another research suggests that a small goggle size is the most significant factor associated with elevated IOP. Therefore, modifying the size of the goggles through increasing the internal diameter of the goggle and merging the two lenses into one can prevent goggle-induced elevation in IOP. Secondly, the use of contrast-enhancing goggle lenses with copper, rose, or brown tints can be particularly helpful through increasing color-contrast or distinction for individuals with low visual acuity and fighting glaucoma-related light sensitivity. This reduces pain and headaches associated with light or glare exposure. It can also be beneficial to add prescription lenses to further improve focus and visual acuity for individuals with Glaucoma. Implementation of a small IOP measuring and sensing device is especially helpful for individuals with Glaucoma and can reduce exercise-induced injury. Similar technology to a non-contact tonometer like Diaton transpalpebral tonometer, which measures IOP through the upper eyelid, can be used to monitor IOP throughout the exercise and create a vibrational sensory response as the result of high IOP so the individual can stop exercise immediately, remove the goggle or move into an upright position to lower their IOP and prevent serious injury. The resulting data from the device can also be monitored and recorded in its own distinct application, which can be installed on devices like smartphones or watches. This data collection can give an insight into the exercise-intensity and positions or types of aquatic activities that lead to elevated IOP so the individual can modify or decrease levels of engagement in such intensities or activities. As a result, modification of sport-equipment like swimming goggles plays a significant role in the engagement of individuals with visual impairments like Glaucoma in swimming and other types of physical activity. Minor adjustments to the size and color of the lens and implantation of an IOP measuring and sensing device helps individuals with Glaucoma reap the benefits associated with the aerobic exercise of swimming safely and enjoyably.

Parisa Abedi Khoozani
Session: VVF 9 (Wednesday, June 16th, 11:15am)

Mechanisms for integrating allocentric and egocentric information for goal-directed movements: a neural network approach

Numerous studies have shown that allocentric and egocentric information are combined for goal-directed movements toward visual targets. Specifically, behavioural studies suggested that this combination follows Bayesian integration, e.g. decreasing the stability of visual landmarks decreased the reliance on allocentric information (Byrne et al., 2010; Klinghammer et al., 2017). Additionally, neural recordings found landmark-specific coding in supplementary and frontal eye fields when monkeys performed saccades toward a remembered visual target (Schütz et al., 2020; Bharmauria et al., 2020). Nevertheless, the intrinsic coordinate representation and underlying processes for this combination remain a puzzle, mainly due to inadequacy of current theoretical models to explain data at different levels (i.e. behaviour, single neuron and distributed networks). In response, we aim to build a theoretical framework to tackle this challenge. In particular, we propose a physiologically inspired neural network with two major components. First, a Convolutional Neural Network (CNN) is used to extract the allocentric and target information: Our CNN performs repeated (2 layers) convolution, rectifications, and normalization to first extract low-level features from input images. With the features extracted, the challenge is to combine them to generate an abstract representation (allocentric cues and target position). We address this challenge by training a feature pooling layer at the end of our CNN network. Second, a Multi-Layer Perceptron network (MLP) is used to combine allocentric (extracted feature maps from the CNN) and egocentric information (initial gaze position fed to the network as an additional input): Our MLP consists of three fully connected hidden layers. These three layers incrementally transform the allocentric and egocentric representations into an integrated motor response. Finally, following the MLP, an additional layer transforms motor responses into final gaze positions. We were able to train these physiologically inspired networks to achieve good correspondence with our dataset (MLP: R2 = 0.80). Additionally, the activity of hidden layer units in the MLP was similar to experimentally recorded neural response fields in monkeys (Bharmauria et al. 2020): neuron's coded gaze information at the motor output layer and the spatial coding was partially shifted toward the shifted target position. These results suggest that our framework provides a suitable tool to study the underlying mechanisms of allocentric and egocentric integration.

John Jong-Jin Kim
Session: IRTG 8 (Thursday, June 17th, 9:30am)

Can you place your phone in a picture? Determining the distance and size of an object in a 2D representation of a scene

Introduction: When viewing an object in a 2D representation of a scene such as on a TV screen, people need to rely on monocular cues such as perspective and occlusion to determine its distance. The size-distance invariance hypothesis suggests that the distance of an object of known size may be determined by its retinal size. An object's distance can also be judged relative to other objects in the scene if their physical sizes are known. How well are we able to use the size cue to determine distance when viewing a familiar object when it is embedded in a 2D representation of a scene?
Methods: Using online testing via PAVLOVIA, participants viewed an upright rectangular block (target) in a hallway scene that provided ambiguous distance information. The target was a representation of an object of very familiar size - their own smartphone - the physical size of which was incorporated into the display program. The screen was viewed at ~50cm. Exp1 (N=71): In the first task, participants adjusted the vertical position of the block to match its perceived location in the scene using its visual size. Then they adjusted its size based on the position which they had chosen in the first task. Exp2 (N=56): the order of the tasks was reversed from exp1. First, they adjusted the block's size to be appropriate for its distance (which was also told to them in meters: 4, 8, 12, 16 or 20m), they subsequently adjusted its position using the size they had chosen in the first task. Exp3 (N=44): the tasks and the order were identical to exp2, but target distances were not given. Instead, the scene contained familiar objects (a bicycle and a door) to provide reference sizes.
Results: Overall, participants could differentiate the distances and sizes of the targets in the correct order. When they matched target size to target positions that they had previously set, they made the targets significantly larger: exp1 ($p < .001$). However, when they matched target positions to target sizes (that they had previously set), the new positions were not significantly different from the original positions: exp2 ($p = .632$) and exp3 ($p = .080$).Target sizes adjusted based on position was not affected by whether the scene had familiar objects or not: exp2 vs. exp3 ($p = .096$), or whether the familiar objects were placed at near (8m) or far (16m) distances in the scene: exp3 ($p = .120$).
Discussion: Our results suggest that people can differentiate visual size from the position of the object in a 2D rendering of a scene and can determine its position from its size. However, they set the visual size much too big in general suggesting they were unable to immerse the object into the scene. Being told how far away the object is, or having familiar objects, such as a bicycle or a door, did not affect their judgment. Possible explanations for these findings and suggestions for future studies will be discussed.

Maria Koshkina
Session: VVF 9 (Wednesday, June 16th, 11:15am)

Towards Sports Video Understanding: Player Detection, Classification and Tracking

We explore approaches for player detection, classification, and long-term tracking in team sports. We formulate player classification according to their team as an unsupervised learning problem where jersey colours are not known a priori. We adopt a contrastive learning approach in which an embedding network learns to maximize the distance between representations of players on different teams relative to players on the same team, in a purely unsupervised fashion, without any labelled data. We evaluate the approach using a new hockey dataset and find that it outperforms prior unsupervised approaches by a substantial margin, particularly for real-time application when only a small number of frames are available for unsupervised learning before team assignments must be made. Remarkably, we show that our contrastive method achieves 94% accuracy after unsupervised training on only a single frame, with accuracy rising to 97% within 500 frames (17 seconds of game time). We propose to incorporate team affiliation classification into long-term tracking of players. To this end, we present a novel hockey player tracking dataset and evaluate current state-of-the-art trackers on it. We discuss approaches to address the challenge of tracking players during active play.

Fengbo Lan
Session: VV 6 (Tuesday, June 15th, 11:15am)

Joint Demosaicking / Rectification of Fisheye Camera Images using Multi-color Graph Laplacian Regularization

To compose one 360 degrees image from multiple viewpoint images taken from different fisheye cameras on a rig for viewing on a head-mounted display (HMD), a conventional processing pipeline first performs demosaicking on each fisheye camera's Bayer-patterned grid, then translates demosaicked pixels from the camera grid to a rectified image grid. By performing two image interpolation steps in sequence, interpolation errors can accumulate, and acquisition noise in each captured pixel can pollute its neighbors, resulting in correlated noise. In this work, a joint processing framework is proposed that performs demosaicking and grid-to-grid mapping simultaneously — thus limiting noise pollution to one interpolation. Specifically, a reverse mapping function is first obtained from a regular on-grid location in the rectified image to an irregular off-grid location in the camera's Bayer-patterned image. For each pair of adjacent pixels in the rectified grid, its gradient is estimated using the pair's neighboring pixel gradients in three colors in the Bayer-patterned grid. A similarity graph is constructed based on the estimated gradients, and pixels are interpolated in the rectified grid directly via graph Laplacian regularization (GLR). To establish ground truth for objective testing, a large dataset

containing pairs of simulated images both in the fisheye camera grid and the rectified image grid is built. Experiments show that the proposed joint demosaicking / rectification method outperforms competing schemes that execute demosaicking and rectification in sequence in both objective and subjective measures.

Matt Laporte
Session: IRTG 6 (Wednesday, June 16th, 9:30am)

Robust versus optimal control of movements by neural networks

Optimal feedback control (Todorov & Jordan, 2002) is a popular model of motor behavior. Formally, the optimality of a movement is decided by a function describing a task objective. How well the controller can meet this objective depends on the relevant dynamics of the environment being captured by the controller's model. In classic control theory, this involves an engineer explicitly adding the appropriate terms to the model. For agents that acquire implicit models of the world through a process of learning, as humans do, their models are presumed incomplete. The agent will continue to experience disturbances it cannot (yet) fully anticipate. Sometimes, the agent may learn to anticipate an unspecified disturbance from the environment but be unable to acquire a particular model of the disturbance. In that case, the strategy of optimal feedback control (with an incomplete model) may be inefficient. Robust control is an alternative strategy that admits uncertainty in the specification of the model and prepares for the worst possible disturbance. Typically, a robust controller will output control signals that are amplified compared to those of a respectively implemented OFC. That is, the robust controller increases the forcefulness of movements, and thus their robustness to non-specific perturbations. But such movements are more energetically costly. These opposing costs—spending more energy, versus neglecting surprise disturbances—suggests we should expect a general tradeoff in learning agents between model-based and uncertainty-based (or "model-free") strategies: "efficient where sufficient, robust where I must". Crevecoeur et al. (2019) found that, in a planar reaching task, humans appeared to trade off robust and efficient strategies: increasing their control gains (as evidenced by hand velocities and EMG) trial-by-trial in response to recent, unexpected perturbation trials, and reducing them again the longer the perturbation trials were withheld. However, it remains unknown how these different strategies are implemented, not explicitly by engineered controllers, but implicitly by neural networks, such as in the brain. What are the structural and dynamical consequences to networks trained to enact these different strategies? Here, we propose to address this question by training separate RNNs to act like robust and optimal controllers in performing reaching movements. We will dissect the trained networks and compare single-unit and population activities. In particular, we will

investigate the networks' responses to perturbation trials. This work may provide insight into how the brain controls and learns movements. It will also inform a planned subsequent study on the emergence of the efficient-robust tradeoff in single networks trained by reinforcement to perform reaches under shifting environmental dynamics.

References:

1. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. Nature Neuroscience 5, 1226–1235 (2002).
2. Crevecoeur, F., Scott, S. H. & Cluff, T. Robust Control in Human Reaching Movements: A Model-Free Strategy to Compensate for Unpredictable Disturbances. J. Neurosci. 39, 8135–8148 (2019)

Ewen Lavoie

Session: VVF 6 (Tuesday, June 15th, 11:15am)

Embodiment, Visual Attention, and Movement in Real and Virtual Worlds

We translated an object interaction task from the real-world (Lavoie et al., 2018) into VR so that we could dissociate a movement from its visual appearance. Participants completed at least 20 trials in two conditions: Controllers - where they saw a visual representation of the VR controller, and Arms - where they saw a set of virtual limbs, while we tracked their eye and body movements. We recently published in Neuroscience of Consciousness (Lavoie & Chapman, 2021) that participants seeing Arms moved more awkwardly in order to make the virtual limbs look similar to how they would if they were interacting with a real-world object. These movement changes were accompanied by an increase in self-reported feelings of ownership over the limbs as compared to the controllers. And, most surprisingly, we found a correlation between movement changes and ownership over the limbs. To emphasize, we found that the participants who moved more differently between the two conditions were also those who reported greater feelings of ownership to the virtual limbs. Overall, this suggests our movements are planned to provide optimal visual feedback, even at the cost of less efficient movements. There is something about seeing a set of limbs in front of you, doing your actions, that affects your moving, and in essence, your thinking. In a second study, we compare the similarities and differences in eye-hand coordination between the virtual version of this task and its real-world counterpart, providing the first detailed analysis of how “real” virtual reality actually is.

Keyi Liu

Session: VVF 11 (Thursday, June 17th, 11:15am)

Instance Segmentation with Sparse Distance Transform Map

Principal component shape representations have proven effective for single-stage object instance segmentation. However, neurophysiological studies suggest that intermediate shape representations

in primate cortex are sparse. Here we explore whether a sparse shape coding strategy can lead to more accurate single-stage object instance segmentation. Specifically, we encode ground-truth 2D shapes using a distance transform and then learn a sparse basis in this distance transform representation. We then design and train a single-stage segmentation head to estimate the sparse

coefficients representing the shape of each object instance, which can be recovered from the zero-crossing level set of the corresponding distance transform map. We demonstrate that this sparse shape encoding method leads to gains over baselines on the MSCOCO benchmark. These results suggest that the sparse shape representations observed in primate cortex can be useful for computer vision object detection systems.

Gaelle Luabeya
Session: VVF 3 (Monday, June 14th, 11:15am)

Instance Segmentation with Sparse Distance Transform Map

Goal-directed arm movements are essential for effective interactions with the external world. They require the integration of multiple sensory modalities to monitor the various muscles involved in the action. Goal-driven actions are generally studied in prehension, a specific movement type used to reach and grasp objects. In prehension, the reach component needs to be coordinated with the grasp component in order to effectively acquire the object. Neural imaging studies have revealed that two streams in the dorsal pathway are running in parallel; the dorsomedial parieto-frontal pathway carries information relevant to the reach component, and the dorsolateral parieto-frontal pathway carries information relevant to the grasp component. The neural correlates for planning reach-grasp integration are not yet established. Previous studies have provided clues on the possible areas involved with these two streams, which might monitor such an integration. However, none of these studies were conducted to investigate the reach-grasp integration directly. In order to answer this question, a functional magnetic resonance imaging (fMRI) experiment was conducted, and we used a cue-separation paradigm to identify the brain areas involved in the integration of reach and grasp during movement planning. Twelve participants were asked to grasp vertically or horizontally a cubic object presented to their left or to their right for this task. Their grasping movement onset was preceded by two successive cues, either a visual cue of the target location or an auditory cue of the target orientation. Each cue presentation was followed by a delay period. Whereas the first delay period only required participants to remember one cue, the second delay period required participants to remember two cues and integrate them as they prepare to initiate the reach-and-grasp movement. We conducted univariate analysis to identify areas that enhanced activity during the integration process. Preliminary results from a conjunction analysis revealed significant activities in the Dorsal Premotor Cortex, Supplementary Motor Area, Cerebellum, and Primary Somatosensory Area. Therefore, the integration of reach location

with grasp orientation does not occur in separate, special-purpose brain areas but rather within well-known motor-related areas.

Paria Mehrani
Session: VVF 4 (Tuesday, June 15th, 9:30am)

Learning a sparse representation in V4 cells reveals shape encoding mechanisms in the brain

The mechanisms of local shape information transformation from V1 to more abstract representations in IT are unknown. Studying the selectivities in intermediate stages of transformation suggest plausible mechanisms. For example, Pasupathy and Conner [1] studied Macaque V4 responses to convexities and concavities. They found that these neurons are selective to complex boundary configurations at a specific position in the stimulus, for example, a convexity adjacent to a broad concavity. Although such investigations reveal intermediate shape representations in the brain, they often do not suffice in capturing complex and long-range interactions within the receptive field due to imposing priors on tunings, e.g., fitting a single Gaussian to neuron responses. Here, we propose a learning-based approach that eliminates the need for such strong priors. Specifically, we investigate shape representation in Macaque V4 cells and formulate shape tuning as a sparse coding problem according to previous findings of V4 neurons [1]. We emphasize that our goal is not to find a mapping from the stimulus to V4 responses but rather to study how V4 neurons combine responses of V2 neurons, in this case, curvature-selective V2 cells, to achieve their reported part-based selectivities. To this end, our proposed model takes responses of simulated curvature-selective V2 cells as input by combining two previously introduced hierarchical models 2DSIL [2] and the RBO network [3]. With simulated curvature signal as input, our algorithm learns a sparse mapping to V4 responses that reveals each Macaque V4 cell's tuning and the mechanism by which the tuning is achieved. Our model captures sophisticated interactions within the receptive field from neuron responses. Our results on V4 shape representations confirm long-range interactions between components of a larger shape, providing a better understanding of shape encoding mechanisms in the brain. Acknowledgements We perform our experiments on V4 cell responses that were provided by Dr. Anitha Pasupathy for which we are grateful. This research was supported by several sources for which the authors are grateful: Air Force Office of Scientific Research [grant number FA9550-18-1-0054]; the Canada Research Chairs Program [grant number 950-231659]; and the Natural Sciences and Engineering Research Council of Canada [grant number RGPIN-2016-05352].

References:

1. Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. Journal of neurophysiology, 86(5), 2505-2519.
2. Rodríguez-Sánchez, A. J., & Tsotsos, J. K. (2012). The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. PLoS One, 7(8), e42058.
3. Mehrani, P., & Tsotsos, J. K. (2019). Early recurrence enables figure border ownership. arXiv preprint arXiv:1901.03201.

Laura Mikula
Session: IRTG 2 (Monday, June 14th, 9:30am)

Interpreting and integrating vision for adapting to dynamic environments

We can interact with the world in multiple ways, using our own body or tools, with apparent effortless ease. However, complex mechanisms come into play to control and adapt movements to our ever-changing environment. As a result, dynamic visual cues need to be integrated to select appropriate behavioral actions. In response to unexpected visual feedback, the brain has to update the sensory predictions of the ongoing movement and adjust motor commands accordingly to correct for errors. The overall goal of my project is to better understand how the brain interprets dynamic visual information during motor tasks to predict, adapt and control action. Previous studies about sensorimotor adaptation have focused on how people adjust their movements following visual manipulation of the motion of their hands. In contrast, others have investigated the effect of manipulating the visual feedback of action-relevant elements in the world. In virtual and online environments, it is possible to manipulate the physics of familiar objects through perturbations of visual feedback. It is of particular interest to understand how we adapt to the modification of such strong priors, which develop though sometimes years of experience. For this purpose, I am conducting an online experiment consisting of a pong-like game in which a ball is bouncing off a horizontal wall located at the top of the screen. Participants have to intercept the ball by moving a paddle, controlled by their cursor, along the left-right direction. When bouncing back from the top wall, the ball can follow a normal path, or an angular deviation can be introduced to create a discrepancy between the expected and the actual ball trajectory. We expect more and/or larger intercepting errors in the condition where the angular deviation is present. However, the top wall can also be displayed tilted so that it appears consistent with the deviation of the ball from the expected trajectory. The tilted wall gives participants a visual cue to predict the otherwise unusual ball path. This should result in reduced intercepting errors. The experiment is in ongoing development. The goal is to investigate whether participants can adapt to this new paradigm and if so, what is the time course of this adaptation? Furthermore, this study should give some insight as to whether meaningful visual cues are taken into account when a perturbation is introduced in the environment we interact with. Future directions might include examining how this newly acquired visuomotor mapping may transfer to the opposite hand, other tools, or environments.

Krista Mitchnick
Session: VVF 4 (Tuesday, June 15th, 9:30am)

Damage to the hippocampal dentate gyrus impairs the perceptual discrimination of complex, novel objects

The hippocampus (HPC) and dentate gyrus (DG) subregion in particular, is purported to be a pattern separator, orthogonally representing similar information so that distinct memories may be formed. The HPC has also been shown to be involved in fine-grained perceptual discrimination. This role may be limited to spatial/scene stimuli, whereas other medial temporal lobe structures are necessary for the

discrimination of objects, and it is unclear if the DG itself contributes to pattern separation beyond memory. A unique individual, BL, who has bilateral HPC damage that is relatively limited to the DG, was previously shown to have poor discrimination of similar, everyday objects in memory (Mnemonic Similarities Task). Here we demonstrate that BL's deficit extends to fine-grained perceptual discrimination of novel objects. Specifically, BL was presented with matched possible and impossible objects, which give rise to fundamentally different 3D perceptual representations, despite being visually similar. BL performed significantly worse than controls when asked to select an odd object (e.g. impossible) amongst three identical counter part objects(e.g., possible)presented at different rotations(Oddity Judgement). In contrast, BL's performance was indistinguishable from that of controls on a series of other tasks involving the same objects, indicating that he could visually differentiate the object pairs under less demanding conditions (Same-Different Judgement), that he perceived the objects holistically (Depth Judgement), and that he could categorize the objects (Possible and Impossible Judgement). Furthermore, his performance on standardized neuropsychological measures indicated intact visual-spatial attention as well as visual and auditory working memory. Examination of BL's eye fixation patterns during the Oddity task showed that unlike controls, BL made more eye movement transitions between the important regions of the objects compared to within the objects on correct trials, suggesting a significantly increased effort or need to compare the objects. These results provide direct evidence that the DG is necessary for fine-grained perceptual discrimination of novel objects, indicating that the DG might function as a 'global pattern separator' of a wide range of stimuli within perception, and that its role is not limited to memory.

Shanaathanan Modchalingam
Session: IRTG 7 (Thursday, June 17th, 8:30am)

Factors affecting implicit motor learning

When our intended movements have unintended outcomes, the human motor system can quickly adapt future movements. Motor output is modified in a way such that motor errors, that is, the difference between the expected and the perceived consequences of any motor output, are reduced. Both explicit and implicit neural processes play key roles in adaptation. Explicit processes, such as the use of conscious strategies to counter a perturbation, can be quickly employed and allow for flexibility in rapidly changing conditions. Implicit processes on the other hand, such as the unconscious updating of internal models, arise slowly and allow for reliable, persistent changes in the motor system. When adapting reaching movements, implicit components of adaptation have been demonstrated to have an upper boundary in the amount by which they alter future movements, regardless of the size of the experienced perturbation [1]; that is, they are thought to be limited in scope. We explore methods that may allow larger amounts of implicit learning and show that the proposed upper boundary on implicit adaptation can be overcome. We also show that the additional implicit learning observed is not caused by simply

gradually increasing a perturbation to prevent explicit strategies. This suggests that while additional implicit adaptation is difficult to achieve during tasks with large, salient errors, and even in tasks with slowly introduced errors, it may be achievable by evoking multiple distinct error correction steps. Furthermore, I will introduce future studies where, using immersive virtual environments, we explore whether implicit learning is sensitive to visual changes in the environment. Implicit error corrections are efficient and arise even with limited processing time before the onset of a movement. It is unclear to what extent visual feedback is used to inform such movements in the limited time available. We will try to determine if implicit learning is purely error driven, or if predictable physical interactions in the task environment inform the implicit learning process.

Reference: Bond KM, Taylor JA. Flexible explicit but rigid implicit learning in a visuomotor adaptation task. J Neurophysiol. 2015

Sanjida Sharmin Mohona

Session: VVF 11 (Thursday, June 17th, 11:15am)

Effects of Chromatic Aberration Compensation on Visibility of Compression Artifacts

In virtual and augmented reality displays, lenses focus the near-eye display at a far optical distance and produce a large field of view to immerse the user. These lenses typically exhibit considerable distortion and cause chromatic aberration. These are not apparent to the user because they are typically corrected by pre-processing the image with the opposite distortion before sending it to the display. Such pre-processing involves pre-warping source images with inverse pin-cushion (barrel) distortion to correct for the pin-cushion transform from the display optics with different correction for each colour channel. Most image compression algorithms use a colour space conversion before compression which normally improves compression performance by reducing the degree of correlation between components. However, as lens pre-distortion processing is colour specific the spatial correlation between colour channels is disrupted by this processing; objective analyses suggest that the colour space conversion may not be beneficial under these conditions. Here we used the ISO/IEC 29170-2 flicker protocol that has been adapted for 3D imagery, to evaluate the sensitivity of two state-of-the-art display stream compression algorithms to characteristic distortions resulting from stereoscopic head-mounted display pre-processing which either included normal colour transformations or bypassed them. A set of 10 computer-generated stereoscopic high dynamic range images were tested. Images spanned a wide range of content and were designed to challenge the codecs. The pre-processing workflow involved pre-warping the images, compressing with each codec, and finally de-warping with pin-cushion distortion. De-warping was applied to simulate the distortion from magnifying lenses as all images were viewed on a mirror stereoscope without such lenses. The main image manipulations were the codec used, the compression levels and whether the colour transform was bypassed (bypass-on) or not (bypass-off). Images were compressed at the codec's

respective nominal production level and at each image's estimated limit of visually lossless compression. 60 observers were tested in 3 groups of 10 for both codecs. Overall, we found little sensitivity to these distortions and our results confirmed that bypassing colour transforms in the codec can be significantly beneficial for some images.

Rachel Moreau
Session: VVF 4 (Tuesday, June 15th, 9:30am)

Differential processing of reflection and rotation symmetries in visual textures

Symmetries of various types are prevalent in the natural world. Psychophysical studies show that reflection symmetry (found in faces, bodies) can be detected preattentively, requiring less cognitive resources than other symmetry types such as rotation (found in flower petals, snowflakes). The distinction between symmetry types is important to our understanding of how symmetries contribute to perception of scenes and objects. Visual search has previously been used to probe the distinction between serial and parallel processing of cues to object shape (Enns and Rensink, 1991). Here we employ a visual search paradigm with symmetries that are embedded within regular textures. Our goal is to enhance our understanding of the mechanisms responsible for perceiving symmetries in textures, and differentiate between reflection and rotation symmetry. Based on previous findings, we hypothesize that reflection will elicit more parallel processing than rotation. We conducted two visual search experiments in which participants were presented with regular textures consisting of arrays of tiles containing symmetries. Participants were asked to report the presence of a target tile that had no symmetry and thus disrupted the regularity. We used four different array sizes (total # tiles: 9, 16, 25, 36), and trials with each array size were presented in an interleaved fashion. In Experiment 1, the non-target tiles contained reflection symmetry (N=133), while in Experiment 2 they contained rotation symmetry (N=148). In both experiments we found that accuracy was reduced and RT was increased as array sizes got larger, consistent with serial processing. Importantly, this array size effect was reduced for reflection symmetry relative to rotation symmetry. These results suggest that reflection symmetry elicits more parallel processing than rotation and that the two symmetry types can be differentiated in terms of the mechanisms required for perceiving them.

Marie Mosebach
Session: IRTG 7 (Thursday, June 17th, 8:30am)

Linking signal relevancy and reliability in somatosensory predictions

When moving a limb incoming somatosensory signals are rated as less intense compared to a resting state. This phenomenon, also known as somatosensory suppression, can be explained in the context of an internal forward model that generates sensorimotor predictions based on the motor command. Being able to predict the somatosensory consequences of one's own movements helps to distinguish between self-generated and externally-induced stimulations. Moreover, this mechanism frees capacities to process novel and relevant, while suppressing predictable somatosensory signals (e.g. signals arising from the movement itself). The strength of somatosensory suppression further appears to be modulated by the need to process incoming somatosensory signals in order to successfully accomplish the movement. Here, we examined how the relevancy to process somatosensory information influences suppression and whether it is impacted by the reliability of these signals. Participants had to perform a directed reach to an instructed, but invisible target in front of a screen. Feedback that the target was found was either presented tactilely or visually, and thus the processing of somatosensory signals was either relevant for task completion or not. In addition, we varied the reliability of the tactile feedback between groups by presenting either a strong or weak feedback vibration on the palmar part of the moving index finger. To probe somatosensory suppression, participants received an additional short vibrotactile stimulus with varying intensities on the dorsal part of the moving index finger at the start and to the end of the movement. After each movement, they to indicated whether they felt this stimulus. We calculated detection thresholds for each feedback modality and each reliability group. Somatosensory suppression was generally stronger at the beginning than to the end of the movement. As expected, we found stronger suppression for tactile than for visual feedback. This difference was mainly driven by the reliability of the tactile feedback and did not hold when the feedback stimulus was weak, i.e. less reliable. These results indicate that suppression of predicted somatosensory signals is modulated by the need to process somatosensory information and further interacts with the reliability of incoming somatosensory signals.

Lina Mussa
Session: IRTG 6 (Wednesday, June 16th, 9:30am)

Explicit attention to allocentric visual landmarks improves memory-guided reaching

The presence of an allocentric landmark can have both explicit (instruction-dependent) and implicit influences on reaching performance (Byrne and Crawford 2010; Chen et al., 2011 Klinghammer et al. 2015, 2017). However, it is not known how the instruction itself (to rely either on egocentric versus allocentric cues) influences memory-guided reaching. Here, 13 participants performed a task with two instruction conditions (egocentric vs. allocentric), but with similar sensory and motor conditions. In both conditions, participants fixated gaze near the centre of a display aligned with the right shoulder, and an

LED target briefly appeared (alongside a visual landmark) in one visual field. After a mask/memory delay period, the landmark re-appeared in the same or opposite visual field. In the allocentric condition, participants were instructed remember the initial location of the target relative to the landmark, and to reach relative to the shifted landmark. In the egocentric condition, subjects were instructed to ignore the landmark and point toward the remembered location of the target. To equalize motor aspects (when the landmark shifted opposite), on 50% of the egocentric trials subjects were instructed to anti-point i.e., opposite to the remembered target. When the landmark stayed within the same visual field, the allocentric instruction yielded significantly more accurate pointing than the egocentric instruction, despite identical visual and motor conditions. Likewise, when the landmark shifted to the opposite side, pointing was significantly better following the allocentric instruction (compared to motor-matched antireaches). This was true regardless of whether the data were plotted in allocentric (target relative-to-landmark) or egocentric (target-relative-gaze) coordinates. These results show that in the presence of a visual landmark, memory-guided pointing improves when participants are explicitly instructed to point relative to the landmark. This suggests that explicit attention to a visual landmark better recruits allocentric coding mechanisms that can augment implicit egocentric visuomotor transformations.

Acknowledgements: The Vision: Science to Applications (VISTA) program and the Canada Research Chairs Program.

Veronica Nacher

Session: VVF 11 (Thursday, June 17th, 11:15am)

Visual response fields in DLPFC during a head-unrestrained reach task

Reaching involves a coordinated sequence of gaze, head, and arm movements toward a visual stimulus. Several studies have examined eye-head-hand coordination in the human, but the underlying neural mechanisms, especially those controlling head motion, have not been studied. We addressed this problem by recording single neurons from dorsolateral prefrontal cortex (DLPFC) while two trained monkeys performed a reaching paradigm, that allowed unencumbered head motion and reaching in depth. Animals touched one of three central LEDs (different initial hand locations) at waist level while maintaining gaze on a central fixation dot (with a jitter of 7- 10° from trial to trial) and were rewarded if they touched a target appearing at one of 15 locations in a 40° x 20° (visual angle) array. Preliminary analysis of 271 neurons in both monkeys showed an assortment of target/stimulus, gaze, pre-reach, and reach-related responses in DLPFC. Most neurons could be described as falling into three main groups. ‘Early” neurons increased firing rate during the target presentation and gaze onset. “Early-late” neurons responded in a sustained way from target presentation until reach. Finally, “Late” neurons increased the firing rate during the pre-reach and peak when the monkey reaches. We first tested for gaze, head, and hand gain fields during the different neuronal responses and found in both animals that 38% of target, gaze, pre-reach, and reach-aligned responses were gain modulated by initial hand

position. A small fraction of neurons showed gain fields for initial eye position (4%), and for both initial eye and hand position (6%). After removing the gain field effects, we fitted the residual data against various spatial models and found that “Early” neurons best coded the target (Ts) primarily space-centered during target presentation and, during gaze shifts, preferentially coded displacement of the arm (dA). This Ts-dA transformation occurred 150-200 ms after target onset. “Early-late” neurons coded dA from target presentation until reach and “Late” neurons best coded Ts during pre-reach and arm position in future space (Afs) during reach. A more complete analysis will aim to describe the complete coding and distribution of gaze, head, and reach signals in this region.

Katherine Newman
Session: VVF 1 (Monday, June 14th, 9:30am)

Understanding the influence of a brain derived neurotrophic factor gene polymorphism on brain network integrity and noninvasive brain stimulation techniques

Brain-derived neurotrophic factor (BDNF) is involved in the neurogenesis and maintenance of cells throughout the central nervous system. A single nucleotide polymorphism (SNP) on the BDNF gene – resulting in a Valine to Methionine conversion – is associated with both structural (e.g., fewer dendritic spines) and functional changes in the brain (e.g., reduced functional connectivity (FC) within the default network). Additionally, the BDNF Met allele reduces the efficacy of noninvasive brain stimulation (NIBS) techniques, such as repetitive transcranial magnetic stimulation (rTMS), used to treat a range of clinical conditions (e.g., depression, Schizophrenia). NIBS techniques also show potential to improve non-clinical symptoms (e.g., improve goal-oriented attention). Resting-state functional magnetic resonance imaging (RS-fMRI) was used in combination with intermittent and continuous theta burst rTMS in adult human participants to evaluate the influence of the Met BDNF polymorphism on baseline integrity of large-scale brain networks and the capacity to modulate FC within and between these networks. Saliva samples were analyzed for the BDNF Met allele. Behavioural and RS-fMRI data were collected before and after administering rTMS to a key node of the default network. Met allele carriers had reduced baseline FC in the default network, relative to non-carriers. As well, NIBS protocols did not significantly modulate FC in a BDNF Met homozygous participant. We are collecting additional behavioural data and using open-source neuroimaging databases to further investigate how a polymorphism on the most widely expressed neurotrophin in the human brain relates to functional networks involved in perception and higher-level cognition.
Keywords: functional connectivity, fMRI, BDNF, TMS, default mode network.

Edward Ody
Session: IRTG 2 (Monday, June 14th, 9:30am)

Sensory suppression of self-generated visual and auditory action consequences

Forward model theories of sensorimotor control suggest different processing of self- and externally-generated sensory action consequences. Performing an action creates a prediction for the outcome of that action (known as the efference copy). The efference copy-based prediction is compared to the actual sensory outcome and if the two match, perception and neural activity are attenuated, allowing the system to direct energy towards potentially important external events in the environment. Previous EEG studies have shown that auditory N1 ERPs, thought to represent activity in the auditory cortex, are reduced for tones following a button press compared to when tones are passively listened to. For visual stimuli, the few available studies have shown mixed results with some showing attenuation of self-generated stimuli and others enhancement. These studies typically include an active condition, in which participants press a button to receive a tone or a visual stimulus, and a passive condition in which the outcomes are played back with the same temporal sequence. However, if results are replicable when applying more sophisticated control conditions, which consider tactile and proprioceptive information mirroring finger movements in active conditions remain unknown. Furthermore, visual and auditory modalities are usually not investigated in single experiment making comparisons between modalities difficult and a direct link of neural findings to behavior is usually missing. Therefore, in the present study, we adopted a design, with some adaptations to address methodological and conceptual issues present in previous versions. First, we included an electromagnet-powered button which could automatically move the participant's finger as well as be moved by it and presented all stimuli with a fixed delay after the button press. Second, we studied both visual and auditory stimuli in the same participants to investigative whether similar results would be present across both domains. Third, we included an intensity judgement task which would allow us to compare intensity perception between active and passive movement conditions. Self-generated visual stimuli elicited smaller N1s than passively-generated stimuli over occipital electrodes. There were no significant differences in auditory N1 or perception between active and passive tones. Furthermore, we found no significant differences between active and passive conditions in in the intensity judgment task. The results support previous findings by showing that reduced N1 for self-generated visual stimuli is present when finger movement is matched across active and passive conditions. However, we could not show that N1 suppression represents differences in the perception of self- and externally generated action consequences. The auditory condition may have been confounded by additional noise from the electromagnet. Together, I will provide initial evidence that N1 suppression for visual action outcomes cannot be explained by differences in tactile or proprioceptive information, but rather foster the idea that differences are based on efference copy-based predictions. However, future experiments need to be optimized to demonstrate similar effects for the auditory domain and to better link neural correlate to behavioural performance.

Sarah Park
Session: VVF 7 (Wednesday, June 16th, 9:30am)

Will the Colavita Effect Persist in Online Testing?

The Colavita effect is a multisensory phenomenon where we prioritize visual information over auditory information: when presented with a visual and an auditory stimulus simultaneously, participants report the visual stimulus more often than the auditory stimulus. This robust phenomenon has resisted several experimental manipulations. Recently, many studies have migrated online, which presents a particular challenge for cognitive and perceptual studies due to the need for strict environmental controls. We examined if the Colavita effect would be replicable in an online study. Participants were first asked to adjust the volume of an auditory tone so that it could be heard clearly. They then reported the modality of unimodal and bimodal stimuli. We did not observe a Colavita effect: participants did not respond preferentially to the visual component of audiovisual stimuli. The absence of the Colavita effect is likely attributable to both the change and variation in environment between participants.

Sara Pishdadian
Session: VVF 5 (Tuesday, June 15th, 11:15am)

Mnemonic discrimination and spatial abilities in healthy aging and subjective cognitive decline

This poster will share findings from completed and ongoing studies regarding the sensitivity versus specificity of visual tasks to detecting cognitive, and specifically memory, changes in community residing older adults. The first reported findings will be from published manuscript (Pishdadian et al., 2020; Neuropsychologia) evaluating the sensitivity and specificity of the Mnemonic Similarity Task (MST), a sensitive behavioral measure of mnemonic discrimination. Results showed that Montreal Cognitive Assessment MoCA) overall score, MoCA score without delayed recall score and MoCA status based on cut-off score, all predicted MST lure discrimination performance in 94 older adult participants while MoCA delayed recall score did not. The results support the sensitivity of the MST in detecting

general cognitive decline but call into question the specificity of the MST with respect to memory. The second findings are part of an ongoing project investigating how allocentric spatial abilities, trait autobiographical memory abilities, anxiety, and associative memory abilities differentially predict memory dissatisfaction in older adults (subjective cognitive decline; SCD), a condition that increases the risk of developing dementia.

Hossein Pourmodheji
Session: VV 7 (Wednesday, June 16th, 9:30am)

Multiple Pedestrian Tracking by a Mask R-CNN with Post Processing and an Adaptive Information-Driven Motion Particle Filter Model

Multiple Pedestrian tracking (MPT) system is a crucial component in a wide range of applications in computer vision, such as video surveillance, traffic monitoring, and sports analysis, to name a few. In MPT systems, tracking-by-detection is a popular tracking paradigm with two stages (1) Detection and (2) Tracking [1]. The tracking performance is dependent on the detection quality in the detection stage. Despite efforts to generate accurate and reliable pedestrian detections, it is still a challenging task for researchers to develop a perfect Multiple Pedestrian Detector (MPD) [2]. In the tracking stage, although many algorithms have been developed over the years, the particle filter (PF) based MPT approaches [3] have shown more promise. Traditional PF approaches suffer from the degeneracy problem, wherein after a few iterations, except for a few particles, all the others have negligible weights. We combine novel post-processing steps with the Mask Region Convolutional Neural Network (Mask R-CNN) to identify multiple pedestrians in a given video frame for the detection stage. For post-processing step 1, we calculate the area of the detected bounding boxes and analyze the area distribution in each frame. In each frame, the bounding boxes with an associated area less than the lower area threshold or greater than the upper area threshold will be removed. For post-processing step 2, we propagate the high confidence pedestrian detections from the previous frame and create the final detection set for the current frame [4]. We also propose a robust MPT algorithm using enhanced particle filtering with an adaptive information-driven motion (AIDM) model and resampling scheme for the tracking stage. This algorithm retains information of the highly weighted particles during the particle propagation and resampling steps. It also injects new particles generated from the associated pedestrian detection with the tracker [5]. The proposed algorithm was evaluated on multiple video sequences taken from two publicly available datasets. It achieves superior performance compared to an MPT algorithm implemented using a particle filter with a linear constant velocity motion model (MPT-LCVMPF) and

other state-of-the-art MPT algorithms. Moreover, the tracking accuracy and precision are significantly improved, and the number of tracker identification (ID) switches is reduced simultaneously.

References:

1. G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, “Deep learning in video multi-object tracking: A survey,” Neurocomputing, vol. 381, pp. 61–88, 2020.
2. K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” 2017, pp. 2961–2969.
3. X. Wang, T. Li, S. Sun, and J. M. Corchado, “A survey of recent advances in particle filters and remaining challenges for multitarget tracking,” Sensors, vol. 17, no. 12, p. 2707, 2017.
4. M. Al-Shatnawi, V. Movahedi, A. Asif, and A. An, “Improving Real-Time Pedestrian Detection Using Adaptive Confidence Thresholding and Inter-Frame Correlation,” 2018, pp. 1–5.
5. M. Al-Shatnawi, A. Asif, V. Movahedi, A. An, Y. Hu, and J. Liu, “Multiple pedestrian tracking based on modified mask R-CNN and enhanced particle filter using an adaptive information driven motion model,” 2020, pp. 4–12

Lucie Preißler

Session: IRTG 6 (Wednesday, June 16th, 9:30am)

Emotion perception of interactive body movements in preschoolers

Body movements, similar to faces and speech, convey emotions. Infants are able to recognize emotions from human action kinematics as early as in their first year of life (Addaboo et al., 2019). However, very few studies have investigated the role of social interactions in emotion perception from body movements in childhood. Perceiving such interpersonal communication is important for the understanding of social intentions and social behavior. Moreover, it is crucial for the ability to anticipate how an action unfolds. So far, research has shown that preschoolers are able to recognize social actions from movement kinematics (Zhai et al., 2019) and that school-aged children are more accurate in action perception from interactive dyads compared to monads (Ghanouni et al, 2005). However, it is yet to be determined whether this is also true for the perception of emotional content in interactive stimuli. The aim of the study is to find out how preschoolers perceive the emotional content of actions from movement kinematics. Therefore, we presented 5-year-old children with point-light displays (PLD) of emotional body movements (happy, angry) in versions of (i) the original pairs (dyads), (ii) a single actor (monad), and (iii) scrambled dyads. In condition 1, children were asked to categorize the emotions via button press while reaction times were measured. In condition 2, we recorded facial electromyographic (EMG) activation of the muscles involved in happy (zygomaticus major, ZM), angry (corrugator supercilii, CS) and fearful (frontalis, F) facial expressions during PLD observation. Each participant completed both conditions to allow for conclusions about conscious and unconscious

emotion perception. We expect emotion-specific facial muscle activity in response to the emotional PLD, and a higher emotion recognition accuracy for interactive dyads compared to monads. In addition, we expect faster processing of happy compared to angry PLD, as such a positivity bias has also been found for the processing of emotional facial expressions and emotion terms (Kauschke et al., 2019).

Hamid Ramezanpour
Session: IRTG 2 (Monday, June 14th, 9:30am)

Temporal cortex participates in attention control

Attention is an indispensable component of active vision. Contrary to the widely accepted notion that temporal cortex processing primarily focusses on passive object recognition, a series of very recent studies emphasize the role of temporal cortex structures, specifically the superior temporal sulcus (STS) and inferotemporal (IT) cortex, in guiding attention and implementing cognitive programs relevant for behavioral tasks. The goal of this paper is to advance the hypothesis that the temporal cortex, classically thought to be only involved in object recognition, entails necessary components to participate in attentional control in a flexible task-dependent manner. First, we will briefly discuss the general architecture of the temporal cortex with a focus on the STS and IT cortex and their modulation with attention. Then we will review evidence from behavioral and neurophysiological studies that support their guidance of attention in the presence of cognitive control signals. Next, we propose a mechanistic framework for executive control of attention in the temporal cortex. Finally, we summarize the role of temporal cortex in implementing cognitive programs and discuss how it might contribute to the dynamic nature of visual attention to ensure flexible behavior.

Renate Reisenegger
Session: IRTG 5 (Wednesday, June 16th, 8:30am)

Neurophysiological correlates of self-motion processing and navigation

Introduction: The processing of visual self-motion information is an integral part of spatial navigation. Numerous studies demonstrated the involvement of the macaque medial superior temporal area (area MST) and ventral intraparietal area (area VIP) in the processing of visually simulated (and real) self-motion. Functionally equivalent areas haven been identified in humans by means of fMRI. In the past years, further areas along the visual cortical hierarchy, such as area 7a and the primary visual cortex (V1), have also been implicated in self-motion processing. Their specific functional role, however, is far from being understood. Likewise, it is still unknown, how information about self-motion is integrated with spatial navigation in the brain. Over the last decades, several studies have shown evidence for a causal involvement of the hippocampus and the parahippocampal, entorhinal and retrosplenial cortices

in spatial navigation, scene detection and spatial memory. Different classes of neurons are considered to contribute to establishing a cognitive map: a representation of the spatial environment to support locating oneself and the guidance of future navigational action. Curiously, there has been a “mutual exclusion” of research referring to the (para-)hippocampal formation and related brain regions and their role for spatial navigation and work investigating self-motion responses at the visual cortical level. As an example, little is known about time-resolved interactions between the hippocampus and visual cortical regions during (visually simulated) self-motion and path integration. Importantly, the electrophysiological signature of hippocampal-visual cortical interactions during path integration in humans has not been investigated.

Methods: Our aim is to bridge the gap in understanding of the functional relationship between the neural self-motion- and the spatial navigation-network by studying the interaction between hippocampus and visual cortical regions during path integration. We will perform intracranial EEG (iEEG) recordings from presurgical epileptic patients. Patients will solve a distance reproduction task while their brain activity is continuously recorded through electrodes implanted for clinical purposes. More specifically, patients will be presented a visually simulated self-motion across a ground plane (passive condition). Next, they will be asked to reproduce double the observed distance (active condition). Speed profiles will be recorded and presented to them after a block of active and passive trials (replay condition). Data from each trial from the active and the respective replay conditions will be compared at two points: 1) the subjective single distance, i.e., half of the traveled active distance; and 2) the objective single distance, i.e., the travel distance of the passive displacement, which had to be reproduced twofold. Data will be analyzed with a continuous complex Morlet-wavelet transformation followed by cluster-based permutation analyses.

Hypothesized results: Firstly, we expect to find reduced responses to self-motion onset in the passive as compared to the active condition. Secondly, we expect enhanced hippocampal-driven cortical activity when passing half of the reproduced path, i.e., a neural correlate of subjective distance, especially in the theta band

Alica Rogojin

Session: VVF 5 (Tuesday, June 15th, 11:15am)

White matter microstructure changes are associated with declines in cognitive-motor task performance in older adults with a genetic (APOE e4) risk for Alzheimer’s disease

Introduction: Cognitive-motor integration (CMI) involves concurrent thought and action, which requires the interaction of large networks in the brain. Previous findings have shown that CMI performance is impaired in individuals with specific dementia risk factors (family history of dementia and presence of the APOE e4 allele) in advance of any cognitive impairments. These findings suggest that CMI impairments are associated with early dementia-related brain changes. The objectives of the current

research study are to assess the relationship between white matter integrity and cognitive-motor performance, and the impacts of dementia family history, sex, and APOE status.
Methods: Participants included right-handed older adults with a high-risk (n=25, 12 female) and a low-risk (n=24, 12 female) for dementia. Participants were tested on four visuomotor tasks where reach and gaze were increasingly spatially dissociated using two linked touchscreens. These tasks included a standard condition requiring direct interaction with visual targets, and three dissociated non-standard conditions requiring CMI (visual feedback reversal, plane-change, and plane-change + feedback reversal). APOE genotyping was determined from salivary measures, and diffusion tensor imaging (DTI) measures of fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity were collected to look at white matter integrity.
Results: Changes in DTI measures in several major white matter tracts were a significant predictor of poorer CMI performance in individuals with the APOE e4 allele. These data support our hypothesis that disruptions in CMI performance are associated with identifiable brain alterations early in disease progression.
Conclusion: The preliminary findings provide insight into the impact of AD pathology on neural networks underlying complex visuomotor transformations and demonstrate that the CMI paradigm discussed in this study may be used as a non-invasive, easily accessible assessment tool for dementia risk.

Lisa Rosenblum
Session: IRTG 2 (Monday, June 14th, 9:30am)

Influence of tactile flow on visual heading perception

Self-motion through an environment induces sensory signals in various sensory modalities: visual, vestibular, tactile and even olfactory, which help to determine parameters like travelled distance (path-integration) or direction (heading). While the processing of visual-vestibular self-motion information is subject to a growing body of literature, the processing of visuo-tactile signals in the context of self-motion has received comparatively little attention. Here, we investigated the interplay of visual and tactile heading cues in the perception self-motion direction in a set of behavioral studies. In a first experiment, participants were presented visual optic flow stimuli on a large tangent screen (81° x 66°) simulating self-motion across a ground-plane in one of nine directions (from-16° to +16° in 4° steps). Tactile stimuli were delivered by air flows from head-mounted nozzles around the participants' head (from -24° to 24° in 8° steps. In separate blocks of trials, we presented uni-modal (visual or tactile) or bi-modal stimuli. In bi-modal trials, the tactile stimulus was behaviorally irrelevant. Its direction was either congruent with the visual heading direction or had an offset of up 40°. Subjects (N=10) indicated perceived heading in the unimodal, and visually perceived heading in the bimodal trials by placing a mouse pointer on the screen. The second experiment was similar to the first, yet (1) tactile flow was

provided by hair dryers and (2) in the bi-modal condition, the tactile stimulus was behaviorally relevant. In the unimodal and the bimodal conditions, participants (N=10) had to indicate perceived heading. Self-motion directions in the bimodal condition were either congruent or had an offset of the tactile flow of up to 30° with respect to visual heading. In both experiments, subjects had to fixate a central target during stimulus presentation (500 ms). In the first experiment we found a small but systematic influence of task-irrelevant tactile flow on visually perceived headings as a function of the directional offset. We observed the strongest effect for an offset of 12° between heading stimuli of both modalities. When tactile flow became behaviorally relevant, its influence on perceived heading increased significantly: Perceived heading was shifted more towards the tactile self-motion direction for increasing angular separation between both heading directions. We conclude that tactile flow is more tightly linked to self-motion perception than previously thought. We speculate that this behavioral finding could be linked to multisensory cortical processing stages like the ventral intraparietal area (VIP). Research on the animal model, i.e. the macaque monkey, has shown the position of the receptive fields of neurons in area VIP and their preferred motion direction to be congruent for visual and tactile stimuli. (Bremmer et al. 2002; Avillac et al. 2005). Importantly, a functional equivalent of macaque area VIP has been identified in humans (Bremmer et al. 2001).

Acknowledgements: This work was supported by Deutsche Forschungsgemeinschaft: CRC/TRR135 (project-#: 222641018) and IRTG-1901-The Brain in Action.

Tiina Rosenqvist

Session: VVF 7 (Wednesday, June 16th, 9:30am)

What are Color Visual Systems Doing?

In this project I sketch a novel philosophical theory of color perception and show how it can explain and accommodate a wide variety of perceptual phenomena. In short, I suggest that color perception is “competence-embedded.” By this I mean (i) that the biological function of color vision is to enable and enhance the manifestation of perceptual competences such as scene segmentation, figure-ground segregation, depth perception, object identification, and re-identification (what should be included in this list is largely an empirical question, and the answers might vary across species) and (ii) that color experiences are correct when they do in fact help animals manifest such competences and incorrect when then fail to do so. The framework allows me to differentiate between two kinds of “good” cases of color perception: ideal cases and non-ideal cases. This in turn suggests a plausible explanation for the difference between ordinary perceptual scenarios and some puzzling perceptual phenomena that have to do with simultaneous contrast and assimilation effects. The basic idea is that color vision is embedded in a network of perceptual competences and that in most ordinary situations the demands of these competences converge (e.g. seeing an apple as red helps with figure-ground segregation, scene segmentation and object identification all at once). These are the ideal cases. When it comes to

cases that give rise to simultaneous color contrast or color assimilation effects, the demands of the relevant competences diverge, but this divergence need not imply failure on the part of the color visual system. In my view, when the color “spreads” in watercolor illusions or when the perceived color of a surface changes dramatically due to contrast effects, the color visual system is still doing what it is supposed to be doing—-i.e., helping us manifest perceptual competences. This means that the apparent strangeness of many “color illusions” can be explained by the conflicting of demands, without having to attribute failure to the color visual system itself. In other words, many so-called color illusions can be understood as non-ideal good cases of color perception. Finally, in order to accommodate cases like “the Dress” (the viral internet sensation from 2015), I suggest that we allow for degrees of correctness in color perception. A color experience that enables or enhances the manifestation of at least one perceptual competence is minimally correct, but an experience of the same object that plays the enhancement role with regards to more competences is more correct than the first one.

Jennifer Ruttle
Session: VVF 2 (Monday, June 14th, 9:30am)

Implicit components of motor learning saturate faster with full control of movement and visual feedback

People can quickly adapt their movements to various perturbations, which is usually attributed to explicit components. However, it is unknown how quickly implicit components of learning emerge, as this has never been measured directly but has merely been inferred as the residual aspect of explicit learning. Here, we will discuss a series of experiments where we directly measure implicit components of learning, like reach aftereffects and changes in estimates of hand location, following every single training trial when reaching with a 30° visuomotor rotation. In stark contrast to the assumption that the implicit stage of learning is slow, we find that these direct measures of implicit learning asymptote almost immediately. In our first study, reach aftereffects and changes in hand localization hit their usual maximum magnitude within respectively three and one training trials. In our second study we test if and by how much this rapid implicit measure of learning can be slowed. We test both terminal cursor feedback as well as the absence of movement during training (“passive exposure”), so that there are no sensory prediction errors guiding adaptation. But hand localization shifts still saturate within 3-4 trials. Even in our third study, where we change the perturbation size and direction every 12 trials, we see immediate changes in reported hand location. In short, while updated estimates of hand position and reach aftereffects do not reflect explicit strategies to counter a perturbation, they change very quickly and robustly. Our results also challenge the untested assumption that the time course of implicit learning can merely be inferred from that of explicit learning. These two processes likely occur simultaneously and mostly independently.

Harpreet Saini
Session: VVF 7 (Wednesday, June 16th, 9:30am)

Color modulates feature integration

Recent studies have demonstrated the modulatory role of different isoluminant colors on higher-level cognitive processes, like response inhibition. In this study, we investigated the effect of color on the lower-level process of feature integration using the flash-jump illusion. In this illusion, when a moving bar changes color at a single location along its apparent motion trajectory, the color change is mislocalized and misperceived to occur farther along the trajectory. Our results demonstrated that the isoluminant color of the flash modulated the magnitude of the flash-jump illusion such that participants reported more accurate and precise estimations of the flash location for both red and blue flashes, as compared to green or yellow flashes. As a Bayesian perceptual framework is proposed to underlie the flash-jump illusion, our findings suggest that different colors have different Bayesian weights, which then give rise to inherent and automatic color modulations in the visual system.

Ala Salehi
Session: VVF 2 (Monday, June 14th, 11:15am)

Unsupervised Data Mining from Videos for Improved Depth Estimation in Driving Datasets

One of the most challenging computer vision problems is inferring accurate depth information of a sensed scene. It has many applications in autonomous vehicles, augmented reality, and robotics. In this field, stereo is the approved choice to indicate disparity from two or more images, capture the same area from different points of view. However, inferring depth from a single image is appealing because a stereo rig and overcoming some intrinsic limitations of a binocular setup is dispensable. In a real application, while autonomous systems capture video, the main challenge is the large size of the dataset. It is required to propose a method for inferring real-time depth maps from all frames (approximately 30 frames per second). Depth estimation task in some hard regions is complicated for the model for calculation. Accordingly, they decrease the accuracy and efficiency of the depth estimation model. In this work, we propose a monocular-based depth estimation approach in video datasets. To gain high accuracy, first, we show how large numbers of hard negatives can be obtained automatically by analyzing the output of a trained depth estimator on video sequences. In particular, wrong regions in the depth map that are isolated frames in time, which have not been associated with preceding and following frames, are likely to be hard negatives. We describe procedures for mining large numbers of such hard negatives from unlabeled video data. Then the Adabins method is trained on the hard negative dataset, a transformer-based architecture block that divides the depth range into bins whose center value is estimated adaptively per image. Our results demonstrate an enhancement on KITTI video datasets across threshold accuracy and root means squared error; the threshold accuracy improves by 2.7% compared to the baseline model. Finally, our method can be extended for similar video processing problems.

Elef Schellen
Session: IRTG 2 (Monday, June 14th, 9:30am)

Neural Markers of Size Perception in Dynamic Environments

The neural mechanisms of size perception can be studied with EEG by means of the SSVEP technique. This technique relies on the retinotopic properties of the early visual areas, as well as the high temporal resolution with which these areas are updated. Recent research found that this neural representation of objects varies not only with retinal size, but with perceived size as well. In an attempt to find out more about our neural representation of size in a dynamic environment, we conducted a series of experiments in virtual reality. Moving from 2d, stationary stimuli to dynamic, 3d stimuli has presented a series of challenges. Interestingly, it has shown that movement influences the neural representation of an object more than the retinal or perceived size. Additionally, objects are given relatively constant neural representations, and object identity affects visual processing even down to the lowest levels of perception.

Christina Schmitter
Session: IRTG 3 (Tuesday, June 15, 8:30am)

Discrete vs. continuous action feedback: Commonalities and differences in predictive neural processing

Sensory feedback of voluntary actions is highly predictable and thus engages less neural resources compared to externally generated sensory events. While this has frequently been observed to lead to attenuated perceptual sensitivity and suppression of activity in sensory cortices, some studies conversely reported perceptual enhancement for sensory action feedback. These divergent findings might be explained by the type of feedback stimuli, i.e., discrete action-outcomes vs. continuous action feedback. Therefore, we investigated the impact of the type of action feedback (discrete vs. continuous) on perceptual and neural processing. During fMRI data acquisition, participants detected temporal delays (0 – 417ms) between actively or passively generated wrist movements and visual feedback that was either continuously provided during the movement or that appeared as a discrete outcome. Both feedback types resulted in a neural suppression effect (active < passive) in a largely shared network including bilateral visual and somatosensory cortices, cerebellum and temporoparietal areas. Yet, compared to discrete outcomes, processing continuous feedback led to stronger suppression in right superior temporal gyrus, Heschl´s gyrus, and insula suggesting specific suppression of features linked to continuous feedback. Furthermore, BOLD suppression in visual cortex for discrete outcomes was specifically related to perceptual enhancement. Together, these findings indicate that neural representations of discrete and continuous action feedback are similarly suppressed but might depend on different predictive mechanisms, where reduced activation in visual cortex reflects facilitation specifically for discrete outcomes, and predictive processing in STG, Heschl´s gyrus, and insula is particularly relevant for continuous feedback.

Moritz Schubert
Session: IRTG 7 (Thursday, June 17th, 8:30am)

The Bayesian Causal Inference of Body Ownership model relies on unrealistic parameter choices

Body ownership, i.e. the feeling of a body being one's own, is a fundamental part of human consciousness and important to a variety of clinical issues, e.g. depersonalization experiences in schizophrenic patients. Computational models of cognitive phenomena deepen our understanding of them and allow us to make predictions about relevant outcomes (e.g. the probability to experience a body ownership illusion under certain conditions). One attempt at a computational model of body ownership, the Bayesian Causal Inference of Body Ownership model, explains body ownership as inference about the causes of sensory signals. When sensory signals under consideration (e.g. tactile and visual signals) are attributed to a single object (e.g. a rubber hand), then this object is incorporated into the body. The model generates predictions that are compatible with empirical findings, but (in its current form) relies on unimaginably wide priors. We hold that parameters of cognitive models should be plausible within in the context of the human mind and have implemented a version of the model with one of the priors truncated to a realistic range. This new version is able to produce very similar results compared to the original and, we argue, is therefore preferable.

Jakob Schwenk
Session: IRTG 5 (Wednesday, June 16th, 8:30am)

Processing of temporal information in marmoset primary visual cortex

The accurate representation of temporal information is crucial to visual processing, allowing us to interact with a constantly changing external world. Recent studies have employed a reverse-correlation paradigm to map the temporal response function (TRF) to luminance changes in the human EEG. This approach has revealed that early visual cortex represents stimulus information for far longer than previously assumed, in the form of a narrow reverberation in the alpha band (VanRullen & MacDonald, 2012). While this component, the ‘perceptual echo’, has since been linked to a number of perceptual mechanisms, it has not been described at the level of single cells. Here, we investigated how the responses to the same stimulation procedure as in the previous human EEG experiments present in the local field potential (LFP) and single-unit activity in the visual cortex of marmoset monkeys (callithrix jacchus). In two separate experiments, we recorded in primary visual cortex (area V1) from i) animals passively stimulated under anesthesia, and ii) awake animals allowed to view the stimulus freely (active vision). Visual stimulation consisted of a single patch modulated by a specific random luminance

sequence (flattened white noise) on each trial. Our results show that area V1 reliably represents temporal information at frequencies up to 30 Hz in the LFP and spiking activity. Under anesthesia, LFP coupling to the stimulus was broadband, with peaks in the beta and theta ranges. In the awake, behaving animal, frequency bands were sampled more narrowly with two separate components at slightly higher peak frequencies (beta/alpha). When mapped onto retinal space, phase coupling for both components was strong within receptive field (RF) boundaries, and only weakly present when the patch was not covering the RFs. Previous work in human participants had suggested that the neural signature of the echo would present itself as a travelling wave in the EEG. An additional analysis of phase-gradients did not provide evidence for such a propagation of information across the cortex. Our results provide a first, systematic overview of the representation of broadband temporal information at the level of V1. Contrasting previous findings from human EEG studies, we did not observe long-lasting reverberations of single frequency bands in our data. This could indicate that the echo component is not present in the marmoset cortex, or that it is produced rather by non-linear phase dynamics.
Acknowledgements:

Ayoob Shahmoradi
Session VVF 3 (Monday, June 14th, 11:15am)

Representation as representation-as

I argue that: Representation as representation-as (RAR) Necessarily, for any subject S, object o and time t, if S refers to o at t (by thinking of o or perceiving o), then there is a way φ such that S represents o as φ at t. First, I motivate RAR by showing that it is independently plausible. Then I will show that the arguments presented against this view are not sound. Susanna Schellenberg (2018) argues that RAR entails that perception has to have a sentential format while perception could have a map-like or pictorial structure. She also argues that perceptual reference requires discriminating the referent and even if there are cases of representation-as they must be grounded in discrimination. Regarding the former, I argue that RAR does not entail that perception has to have a sentential structure. In fact, it is hard to see how an image of some object o can represent o without representing it as having certain properties--certain color, shape, location, etc. Regarding the latter objection, I argue that even if Schellenberg were right that perceiving o required discriminating it, that would still entail RAR. Jerry Fodor (1981, 2009) and Zenon Pylyshyn (2004, 2007, 2009) have argued that this view leads to an infinite regress. The idea is that if, for any object o, representing o requires representing it as being ψ, for some ψ, then it is plausible to think that for any ψ, representing ψ itself requires representing it as φ, for some distinct φ. But this leads to an infinite regress which makes object representation impossible in the first place. In response, I will show that i) the assumption that for any property ψ, representing ψ requires representing some distinct property φ is not entailed by RAR. Moreover, ii) I will argue that this further assumption is false for independent reasons. Pylyshyn and Storm (1988) argue, based on their multiple object tracking studies, that if visual representation requires representation-as, then subjects

must represent not only objects but also their properties and their changes over time. keep track of their changes over time. They claim that this is not possible due to our computational limitations. I will argue that i) many of the crucial assumptions that this argument relies on are not supported by their empirical work on multiple object tracking. Also, ii) there are other highly plausible accounts of multiple object tracking that are significantly less demanding in terms of the computational resources that subjects need to employ. I will also consider several other arguments by Pylyshyn (2004, 2007, 2009), Burnston and Cohen (2012) against RAR. I show that their arguments at best show that object representation is not reducible to property representation. While I agree that the reducibility claim is implausible, I argue that, contrary to these authors' claims, it is not entailed by RAR.I discuss blindsight seeing, Bálint's syndrome, inattentional blindness, and other neuro-psychological conditions and show that they could not be accounted for if RAR were false.

Gaeun Son
Session: VVF 1 (Monday, June 14th, 9:30am)

Scene wheels: Measuring perception and memory of real-world scenes with a continuous stimulus space

Precisely characterizing mental representations of visual experiences requires careful control of experimental stimuli. Recent work leveraging such stimulus control in continuous report paradigms have led to important insights; however, these findings are constrained to simple visual properties like colour and line orientation. There remains a critical methodological barrier to characterizing perceptual and mnemonic representations of realistic visual experiences. Here, we introduce a novel method to systematically control visual properties of natural scene stimuli. Using generative adversarial networks (GAN), a state-of-art deep learning technique for creating highly realistic synthetic images, we generated scene wheels in which continuously changing visual properties smoothly transition between meaningful realistic scenes. To validate the efficacy of scene wheels, we conducted a memory experiment in which participants reconstructed to-be-remembered scenes from the scene wheels. Reconstruction errors for these scenes resemble error distributions observed in prior studies using simple stimulus properties. We additionally manipulated the radii of the wheels to parametrically control the similarity among the scenes in the wheels. Upon this manipulation, we found clear evidence that participants' memory performance systematically varied with the level of scene similarity. These findings suggest our novel approach to generate scene stimuli using a GAN not only allows for an unprecedented level of stimulus control for complex scene stimuli, but also that the GAN's latent spaces generating our scene wheels reflect fundamental representational spaces important for human scene

perception and memory. Based on this level of control over the scene stimulus space, we expect that findings from using simple stimuli, such as colour wheels, will generalize to photo-realistic scenes, providing key insights into how we perceive and remember the real-world naturalistic environments that serves as the backdrop to our everyday experiences.

Xue Teng
Session: VVF 8 (Wednesday, June 16th, 11:15am)

Depth Perception Under Scaled Motion Parallax in Virtual Reality

With the fast development of virtual reality (VR) and augmented reality (AR), a number of relevant applications have entered our everyday life, e.g. education, gaming, etc. Depth perception is one of the most important factors in smooth interaction in the virtual environment. It is a complex process, which involves the synchronization of visual, vestibular, kinesthetic and other cues for humans perception systems to work properly. Thus, conflicts between these sources of information may significantly disturb human perception of geometric layout. This work focuses on how the conflict between virtual and physical head motion affects depth perception, in particular object shape. Head mounted displays were used for experiments, e.g. Oculus Rift S or Quest in Rift S emulation mode. A fold stimuls was rendered with a convex dihedral angle formed by two irregularly-textured, wall-oriented planes connected at a common vertical edge. In our experiments, gains were varied with a factor of 0.5 to 2, at which we measured the effect on depth perception. The observers were asked to adjust the angle of the fold till the two joined planes appeared perpendicular. Binocular and monocular conditions were introduced to assess the role of stereopsis. To introduce motion parallax, observers were required to sway laterally at 0.5 Hz. Experimental results show that observers have remarkably accurate perception of object motion under our experiment conditions. Gain has limited impact on depth perception. On the other hand, it has to be mentioned that motion and depth perception cannot be directly quantified, while distance is much easier to measure. Therefore, another set of experiments will be designed to measure how distance perception is affected under the same distorted situation.

George Tomou
Session: VVF 9 (Wednesday, June 16th, 11:15am)

Functional connectivity of transsaccadic perception: Evidence from fMRI and graph theory analysis

While many of the brain regions involved in transsaccadic perception have been discovered, the networks comprised of these regions are not well understood. We formed functional brain networks from functional magnetic resonance imaging (fMRI) data collected during a task designed to dissociate

saccade activity from object feature activity and evaluated the network properties between conditions. Participants (N=17) judged whether a stimulus changed either shape or orientation with or without an intervening saccade. Graph theory analysis of BOLD activation from 50 cortical nodes was applied to identify various network properties including clustering coefficient, local efficiency, global efficiency, and small world propensity. Non-parametric permutation t-test indicated significantly higher values for clustering coefficient ($p = .013$), local efficiency ($p = .013$), and global efficiency ($p = .017$) for saccades relative to fixation, suggesting enhanced functional segregation and integration (i.e., speed of information propagation) during saccades; no significant differences in these measures were found for orientation vs. identity conditions, indicating that the analyzed network specializes in transsaccadic perception and is not involved in the discrimination of object orientation and identity. Further, modularity analysis identified the formulation of different sub-networks during fixation compared to saccade. Specifically, modularity analysis identified three fixation sub-networks: a bilateral dorsal sub-network linking areas involved in visuospatial processing and two lateralized ventral sub-networks linking areas involved in object feature processing. Importantly, when saccades were horizontal and placed the stimuli in different visual hemifields, the two lateralized ventral sub-networks became functionally integrated into a single bilateral sub-network. Betweenness centrality was calculated to provide a measure of each node's significance in linking disparate areas of the network during saccades relative to fixation ($p < .05$), and indicated right hemisphere regions PHC2, vPCu, TO1/MT, vTO, SEF, IPS3, and V7/IPS0, and left hemisphere regions VO1, IPS5, IPS2, LO2, and IPS3 as important hub regions during eye-movements. These results provide objective evidence of a ventral and dorsal stream distinction in human perception and show how sub-networks are modified to functionally integrate as required during saccades.

Jonathan Tong
Session: VVF 8 (Wednesday, June 16th, 11:15am)

Curvature detection and discrimination thresholds for parabolic surfaces depend on the direction of curvature

When direction of surface curvature is ambiguous, based on bi-stable or uninformative lighting and texture cues, observers tend to have a strong bias for reporting convexity (bumps) over concavity (dents). This bias is likely rooted in the propensity of real-world objects to be globally convex. Given this strong expectation for convexity, it is uncertain whether an anisotropy exists in the detection and discrimination of surface curvature in convex compared to concave surfaces, especially when stereopsis is the prevalent cue. We address this question by asking observers to detect and discriminate curvature in Voronoi-textured parabolic surfaces, displayed stereoscopically in a VR headset. Both detection and discrimination tasks followed a sequential 2-alternative-forced-choice (2-AFC) paradigm: in curvature detection, observers compared a planar, frontoparallel, reference surface

against a parabolic comparison surface that varied in curvature direction (concave or convex) and magnitude. in curvature discrimination, observers compared a parabolic reference surface (concave or convex) with a standard curvature magnitude against a parabolic comparison surface with the same direction of curvature but varying in magnitude. In either task, observers were asked to report which surface was more curved. Overall, observers had smaller detection (75% correct) thresholds , as well as discrimination thresholds (just noticeable difference, JND), for convex surfaces compared to concave surfaces. Our results demonstrate that observers are more sensitive to both the presence of curvature, and changes in curvature, in convex surfaces compared to concave surfaces. We speculate that this anisotropy may be driven by the prominence of global convexity in natural scenes

Aaron Tucker
Session: VVF 10 (Thursday, June 17th, 9:30am)

Solving the Conflict Between Breathability and Masked Faces within Facial Recognition Technologies

In 2021, the interlocking crises of the Covid-19 pandemic, wildfires exacerbated by climate catastrophe, and systemic racism in law enforcement have led to the exponential rise in the production of masked faces under the biopolitical management of breathable air. However, the introduction of masked faces into facial recognition technologies (FRTs), a previously successful image-based biopolitical tactic, has short-circuited the technologies' abilities to identify and verify those under its gaze. FRT is a tactic within larger biopolitical strategies tied to precarity and debility which, in 2021, is tasked with processing the novel of combination of media that constructs masked faces, including the elemental media of air and breath; FRTs' initial failures related to masked faces speaks to the internal conflict within contemporary biopolitical control that arise during moments of crises, wherein previous protocols and practices clash with new material and media. In response to the competing vectors brought about by the crises of breathability, the American National Institute of Standards and Technology's (NIST) July 2020 report details FRTs' failures as they relate to masked faces in detail, while providing protocol for future solutions. This paper critiques the solution provided within the report wherein the creation and deployment of synthetic facial data relies on the liminal populations of refugee and immigrants for experimental materials, a practice which ultimately targets the very populations being used to improve the technology

Sarah Vollmer
Session: VVF 8 (Wednesday, June 16th, 11:15am)

Forking Paths: The Living History of Mutable Sound

We present a virtual reality (VR) software system and kinaesthetic environment for sound design that leverages embodied interactivity, tactile intelligence, gestural control, and nonlinear editing. Developed with third-wave HCI in mind, particularly towards challenging the binary of tool and experience, Forking Paths embraces tinkering as a graphic score editor and sequencer where notations are painted as 'living' gestural compositions inVR that come alive as curious and evolving sound-emitting creatures. Against the ocular-centric view of VR as another "visual medium", we consider this medium to operate primarily in personal somatic spacetime where tactile intelligence in this immersive space is considered a natural, yet widely unavailable, extension to creative tools. Forking Paths thus engages the sound designer through the visual, the audible, and the tactile, the latter of which is experienced through the integration of customized on-body haptic feedback and is a reflection of the vibratory existence of each sound creature. This emphasis of sound and haptics thus aims for a synthesis of what is valued in the nuance of human kinaesthetic gesture, along with what spaces of virtual dynamics can offer to intensify it. We therefore identify three practical axes central to this design process to be emphasized: Creation (drawing scores), Interaction (playing improvising), and Forking (nonlinear editing). Pursuant to the importance of exploratory expression and tinkering to studio-based practice, Forking Paths is inspired by the hand-drawn animated sounds and visual musics of pioneers including Oram, Fischinger, McLaren and the Whitneys, as well as the rich histories of graphical notations from Cardew to Xenakis, but also the experimentations of augmented gestures and dynamic canvasses of Haeberli, Maeda, and Levin. Our interactive sound design environment tracks the edit history in such a way that affords later experimentation directly with the edits themselves. The painted gestures are stored as a sequence of deltas, any of which can be singled out and forked-creating an alternate gesture. Each artefact of a sound exists along a temporal continuum and is physically manipulatable (plucking, avoiding, extending, doubling etc.) within the space it consumes. Thus, these graphic notations are not passive, but instead express machine agency and computational creativity as they respond to each other and to their environment. Non-linear editing of their transitional histories offers a novel means of interaction with this generative nature of the sequencer.

Ilja Wagner
Session: IRTG 3 (Tuesday, June 15th, 8:30am)

Humans trade-off set-size and discrimination difficulty in a combined visual search and perceptual discrimination task

In daily life, human observers are confronted with an abundance of potential eye movement targets. To efficiently acquire behaviorally-relevant information in a complex, naturalistic environment, decisions have to be made about which stimulus to select as a target for eye movements as well as high-resolution visual processing, and which stimuli to ignore. Previous studies demonstrated that human eye movement behavior and target selection in general are influenced by factors such as visual saliency (Itti & Koch, 2000) and motivational value (Platt & Glimcher, 1999). Furthermore, it has been shown that human observers weight both of those factors optimally, in order to maximize their performance in visual search (Navalpakkam et al., 2010). However, many everyday situations not only require to select the, for a given situation, most valuable eye movement target, but also to locate said target as fast as possible amidst an abundance of irrelevant stimuli. Here, we investigated if human observers weight the discrimination difficulty of two competing targets and the number of corresponding distractors on the screen (set-size) optimally, while choosing a target in a combined visual search and perceptual discrimination task. In each trial of our paradigm, observers saw a search display composed of two targets (one easy, one hard to discriminate) and a variable number of corresponding easy and hard distractors. Observers were instructed that they can, in each trial, freely choose which target they want to search for and choose for the discrimination task. We gave observers 6.5 minutes to solve as many trials as possible and they received a performance-dependent monetary reward after the experiment. We found that observers considered both, the relative difficulty and the relative set-size of easy and hard stimuli, when selecting a target for visual search. Our results rule out that the observer's behavior was guided by simple set-size-dependent (choose the target embedded in the smaller set) and difficulty-dependent (choose the easier to discriminate target) heuristics. Instead, our data was in agreement with an ideal-observer model, according to which observers, in each trial, select the target that yields the highest monetary reward per time, even if the chosen target is embedded in the larger set or is harder to discriminate. We conclude that human observers can dynamically trade-off set-size and discrimination difficulty while selecting a target for visual search. This trade-off enabled observers to maximize monetary reward per time.

Lee Williams
Session: VVF 8 (Wednesday, June 16th, 11:15am)

I am your ghost

I Am Your Ghost is a multi-platform media work that explores anticipatory grief, death and dying through the use of 360 and 2D camera and 3D computer vision. The works in I Am Your Ghost include experimental short films, virtual reality and digital media artworks. Computational and machine vision are engaged as central to the conceptual framework and visual outputs of the stories told through the project, operating as an embodiment of and simultaneous counterpoint to the highly medicalized aspects of end-of-life care. Engaging a co-creative process between three generations in an Armenian family, the project employs a multimodal approach to autoethnographic research, including interviews, performance elements and archival materials to create its visual language and storytelling. I Am Your Ghost works to express and visually represent the liminal space inhabited by both the dying and surviving members of the family. Collapsing the boundaries between the past, present and future, this project interrogates the concurrent experiences of absence and presence, fiction and reality, and ultimately aims to grapple with what is both impossible and possible. Created under the supervision of VISTA core members Mary Bunch, Caitlin Fisher, and Laurence Harris, the project applies computational and machine vision for visual storytelling in media arts

Noa Yaari
Session: VVF 10 (Thursday, June 17th, 9:30)

How I Think Vision When I Paint

Painting is to ask questions about vision while creating something to look at. Therefore, every addition of content through shapes and colours builds up the questions: How do people see? Do we see the same? And what do my viewers feel and think when they see what I painted? In this presentation, I will show some of my paintings and share my thought associated with them. For example, I know that part of the creative process is to hold a historic perspective on my art, that is, to wonder how people in the future will perceive it. I am also aware of the evolving autobiographical aspect of the works, and their possible value in the market. When it comes to approaching it scientifically, methodologically, I try to view my work as if I were someone else, with no specific knowledge, personality, or experience. This presentation is an opportunity to delve into my “scientific” assumptions and practices, as well as to establish a dialogue with other vision practitioners from various disciplines.

Jason Yu
Session: VVF 1 (Monday, June 14th, 9:30am)

Wavelet Flow: Fast Training of High-Resolution Normalizing Flows

Normalizing flows are a class of probabilistic generative models which allow for both fast density computation and efficient sampling and are effective at modelling complex distributions like images. A drawback among current methods is their significant training cost, sometimes requiring months of GPU training time to achieve state-of-the-art results. This paper introduces Wavelet Flow, a multi-scale, normalizing flow architecture based on wavelets. A Wavelet Flow has an explicit representation of signal scale that inherently includes models of lower resolution signals and conditional generation of higher resolution signals, i.e., super resolution. A major advantage of Wavelet Flow is the ability to construct generative models for high resolution data (e.g., 1024 x 1024 images) that are impractical with previous models. Furthermore, Wavelet Flow is competitive with previous normalizing flows in terms of bits per dimension on standard (low resolution) benchmarks while being up to 15x faster to train.

Jingmin Zhou
Session: VVF 1 (Monday, June 14th, 9:30am)

Multi-label Video Categorization

As the volume of digital videos being created and made available online is rapidly increasing, automatic categorization of video is becoming increasingly important for indexing and search purposes, and beneficial to journalistic and editorial processes of identifying and filtering video content. Although video categorization is an active research field, most reported methods are designed for single-label categorization and are not designed for journalistic purposes. In this work, the goal is to categorize videos into labels meaningful for video editorial purposes, such as “news”, "science", “education”, "business", “technology”, etc. A method for multi-label categorization of video sequences is proposed by experimenting with various classifiers and aggregation methods based on frame and video-level visual features. In most related work, the Convolutional Neural Net (CNN) architectures are used for both feature extraction and classification. To understand the effect of features vs. classifiers in the video categorization performance, we compared the performance of multiple approaches: (i) Object probabilities were obtained by running Inception-V3 [1] on sampled frames and used as visual features. Then the frame-level and video-level features were fed into multi-label Random Forest (RF) [2, 3] classifiers. Aggregations were applied at either feature-level or prediction-level to obtain one prediction for the whole video. (ii) The classifier layer of Inception CNN was re-trained for our categories. Similarly, frame-level predictions were aggregated to obtain one prediction for the whole video. (iii) The classifier layers of Sports 3D-CNN [4] and (iv) CNN-RNN [5] networks were retrained for our categories. These networks provide a video-level prediction. We used 19,515 labelled videos obtained from Vubble dataset (available on http://research.vubblepop.com/), randomly split into 14,629 training and 4,886

test samples (75:25). The median number of labels for each video in this dataset is 6. The highest macro-averaged performance on test set (precision: 0.71, recall: 0.55, F1: 0.53) was reached by L-infinity aggregation (max) of the frame-level predictions of the RF classifiers. F1 values of up to 0.88 were reached for categories with higher number of training samples (majority categories).

References:

1. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 06 2016, pp. 2818–2826
2. V. Losing, B. Hammer, and H. Wersing, “Incremental on-line learning: A review and comparison of state of the art algorithms,” Neurocomputing, vol. 275, pp. 1261–1274, 09 2017.
3. A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, “On-line random forests,” in Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009, 11 2009, pp. 1393 – 1400.
4. D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, “Learning spatiotemporal features with 3d convolutional networks,” in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489– 4497.
5. M. Avendi, PyTorch Computer Vision Cookbook. Pckt, 2020

Workshops

Workshop #1 - Panel discussion on non-academic jobs
Monday, June 14th, 2 – 4pm
Zoom link: **https://tinyurl.com/VirtualVisionFutures1**

Join us for a discussion with several recent CVR graduates and York's Career Development Coordinator about the non-academic jobs available to vision scientists, and about planning strategies for transitioning from graduate school to these positions in several different sectors of industry.

About our panelists:
Dr. Raghavender Sahdev (NuPort Robotics Inc)
Dr. Caitlin Mullin (VISTA @ York University)
Dr. Soo Min Kang (Samsung AI);
Dr. Lindsey Fraser (Vpixx Inc.)
Dr. Carolyn Steele (Career Development coordinator, York University)

Workshop #2 – Panel discussion on academic jobs
Tuesday, June 15th, 2 – 4pm
Zoom link: **https://tinyurl.com/VirtualVisionFutures1**

Join CVR faculty for a discussion on applying for post-doctoral fellowship and academic positions across vision science disciplines.

Dr. Gene Cheung (EECS)
Dr. Gene Cheung received his Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, in 2000. He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He was an assistant then associate professor in National Institute of Informatics (NII) in Tokyo, Japan, from 2009 till 2018. He is now an associate professor in York University, Toronto, Canada. His research interests include 3D imaging and graph signal processing.

Dr. Shital Desai (Design)
Dr. Shital Desai is an assistant professor in Design department, School of Arts Media Performance & Design. Her research focusses on addressing Sustainable Development Goals in global health using human centred and speculative design methods. To that extent, she develops inclusive adaptive technologies for older adults, persons with cognitive impairments and children and cocreates governance policies around global health. Prior to joining York University in 2019, Shital was an AGE-WELL research fellow at Dementia Ageing Technology Engagement (DATE) lab (TRI-UHN) and Inclusive Media Design Centre (IMDC) (Ryerson University). She has 20+ years' experience working in Robotics, Health, communications technology and non-destructive testing industry before embarking upon an academic career.

Dr. Caitlin Fisher (Arts)

Dr. Laurence Harris (Psychology)
Laurence Harris received his PhD from Cambridge University in 1979. After post-docs in Durham (UK) and Dalhousie (Canada) he became a lecturer in Physiology at Cardiff University. He moved to York University in Canada in 1990 where currently holds the York Research Chair in Multisensory Integration. His research interest concerns how the different senses are combined to generate our perceptions. Examples include the visual and vestibular system's role in orientation and self-motion perception; vision and hearing's role in localizing events in space and time; and how knowledge of our body affects our perception of stimuli. He is particularly interested in the way these combinations can adapt to changing demands brought about by unusual environments which he creates using various means including virtual reality, the microgravity of space, human centrifuges, and moving rooms.

Dr. Kevin Lande (Philosophy)
Dr. Kevin Lande is an Assistant Professor in the Department of Philosophy at York University, where he is also a member of the Centre for Vision Research and a Core Member of the Vision: Science to Applications (VISTA) program. He received his PhD from UCLA in 2018 and from 2018-2019 was a postdoctoral researcher at the Centre for Philosophical Psychology at the University of Antwerp. His research is in philosophy of perception and philosophy of psychology.

Workshop #3: Impactful research writing
Wednesday, June 16th, 2 – 4pm
Zoom link: **https://tinyurl.com/VirtualVisionFutures1**
This workshop will focus on writing scientific papers, as well as other forms of scientific writing and knowledge dissemination. Questions are welcome.

Workshop leader:

Dr. Gunnar Blohm, Queen's University