Computational Study of Multisensory Gaze-Shift Planning

Mehdi Daemi

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR DEGREE OF

DOCTOR OF PHILOSOPHY

Graduate Program in Biology

York University

Toronto, Ontario

July 2016

Abstract

In response to appearance of multimodal events in the environment, we often make a gaze-shift in order to focus the attention and gather more information. Planning such a gaze-shift involves three stages: 1) to determine the spatial location for the gaze-shift, 2) to find out the time to initiate the gaze-shift, 3) to work out a coordinated eye-head motion to execute the gaze-shift. There have been a large number of experimental investigations to inquire the nature of multisensory and oculomotor information processing in any of these three levels separately. Here in this thesis, we approach this problem as a single executive program and propose computational models for them in a unified framework.

The first spatial problem is viewed as inferring the cause of cross-modal stimuli, whether or not they originate from a common source (chapter 2). We propose an evidence-accumulation decision-making framework, and introduce a spatiotemporal similarity measure as the criterion to choose to integrate the multimodal information or not. The variability of report of sameness, observed in experiments, is replicated as functions of the spatial and temporal patterns of target presentations. To solve the second temporal problem, a model is built upon the first decision-making structure (chapter 3). We introduce an accumulative measure of confidence on the chosen causal structure, as the criterion for initiation of action. We propose that gaze-shift is implemented when this confidence measure reaches a threshold. The experimentally observed variability of reaction time is simulated as functions of spatiotemporal and reliability features of the cross-modal stimuli. The third motor problem is considered to be solved downstream of the two first networks (chapter 4). We propose a kinematic strategy that coordinates eye-in-head and head-on-shoulder movements, in both spatial and temporal dimensions, in order to shift the line of sight towards the inferred position of the goal. The variabilities in contributions of eyes and head movements to gaze-shift are modeled as functions of the retinal error and the initial orientations of eyes and head. The three models should be viewed as parts of a single executive program that integrates perceptual and motor processing across time and space.

Dedications

To my parents, Hosein and Akram

To my siblings, Mohsen and Elaheh and Ali

To all my friends and family

Acknowledgements

I am extremely thankful to my supervisor Dr. Doug Crawford. He trusted me in the beginning, when it maybe didn't make sense to accept an engineer with no experience in neuroscience, and he supported me all the time during the PhD. He is cool.

I am grateful to my supervisory committee Dr. Laurence Harris and Dr. Keith Schneider. They always supported me and provided invaluable insights.

I am thankful to my lab mates Amirsaman Sajad and David Cappadocia. I have had so many scientific discussions with them through the years, which helped me find solutions for the problems I was facing. They have also been amazing friends.

Candidate Contribution

The candidate (Mehdi Daemi) was responsible for conceiving the ideas and developing the models for all three projects. He did all the work for simulations and analyses provided in the papers. All the manuscripts were written by the candidate.

Dr. Laurence Harris provided valuable comments in writing the first ‘causal inference’ paper. His wisdom and edits vastly improved the manuscript.

Dr. J.D. Crawford supervised the three projects. All projects were conceived under his guidance and feedback. For the third kinematic project, he also contributed to the ideas realized by the model. All the manuscripts were considerably improved by his comments and edits.

Contents

1 General Introduction 1
1.1 Review of Multisensory Phenomena Addressed in this Thesis 2
1.1.1 Inference of the cause of the cross-modal stimuli 2
1.1.2 Variability of reaction times for gaze-shifts towards cross-modal stimuli 4
1.1.3 Eye-head coordination for shifting the line of sight 4
1.2 Review of the Neurophysiology of Cortical Networks 5
1.2.1 Hierarchical organization of perceptual networks in posterior cortex 6
1.2.2 Hierarchical organization of executive networks in frontal cortex 9
1.2.3 Attention 11
1.2.4 Working memory 13
1.2.5 Reasoning and Inference 14
1.2.6 Decision making 16
1.3 Head-Free Gaze-Shifts in Three-Dimensional Space 17
1.3.1 Kinematics of eye-head coordination 17
1.3.2 Subcortical networks of eye-head coordination 18
1.4 Review of Computational Neurocognitive Architectures 20
1.4.1 Criteria for evaluation of different architectures 21
1.4.2 Symbolic, connectionist, and dynamicist approaches 23
1.4.3 Reconciliation of different approaches 24
1.4.4 Neural Engineering Framework 25
1.5 Moving from the literature to our computational models 27
2 Causal Inference for Cross-Modal Action Selection: A Computational Study in a Decision Making Framework 30
2.1 Abstract 31
2.2 Introduction 32
2.3 Model Overview 35
2.4 Mathematical Formulation 40
2.4.1 Method 40
2.4.2 Unisensory Signals 41
2.4.3 Short-Term Memory 42
2.4.4 Spatiotemporal Similarity Measure 44

2.4.5 Decision Making Process 45
2.5 Results 48
2.5.1 Inference of a Unique Cause for Cross-Modal Stimuli 48
2.5.2 Effect of Spatial Disparity 51
2.5.3 Effect of Temporal Disparity 53
2.5.4 Effect of Stimulus Reliability 55
2.5.5 Effect of Evidence Accumulation 57
2.6 Discussion 59
3 Reaction Time Variability of Multimodal Gaze-Shifts: A Computational Study in a Decision Making Framework 62
3.1 Abstract 63
3.2 Introduction 64
3.3 Model Overview 66
3.4 Mathematical Formulation 70
3.4.1 Method 70
3.4.2 Decision Making 71
3.4.3 Confidence Measure 73
3.4.4 GO Command 74
3.5 Results 75
3.5.1 Unisensory Situation 75
3.5.2 Effect of Spatial Disparity 77
3.5.3 Effect of Temporal Disparity 79
3.5.4 Inverse Effectiveness 82
3.5.5 Summary of Results 84
3.6 Discussion 86
3.6.1 Implications for theories of multisensory action initiation 86
3.6.2 Significance for interpreting previous behavioral findings 87
3.6.3 Implications for neurophysiology of multisensory processing 88
3.6.4 Conclusion 89
4 A Kinematic Model for 3-D Head-Free Gaze-Shifts 91
4.1 Abstract 92
4.2 Introduction 93

4.2.1 Overview of Gaze Kinematics 93
4.2.2 Gaze Control Models: from 1-D Saccades to 3-D Eye-Head Control 96
4.2.3 Aims of the Current Study 98
4.3 Model Formulation 99
4.3.1 Overview 99
4.3.2 Basic Mathematical Framework 103
4.3.3 Motor Mechanisms of Eye-Head Movement 106
4.3.4 Static Kinematic Model 108
4.3.5 Solving the Static Model 111
4.3.6 Simulation of full Movement Trajectories 115
4.4 Results & Discussion 115
4.4.1 Gaze Accuracy and the 3-D Reference Frame Transformations 116
4.4.2 Eye, Head and Gaze Orientations and their Constraints 118
4.4.3 Development of the Eye, Head and Gaze Orientations during Gaze-Shift 121
4.4.4 Eye-Head Coordination Strategies Influence Eye-in-Space Orientation 125
4.5 Concluding Remarks 130
5 General Discussion 132
5.1 Review of Causal Inference for Multimodal Gaze-Shift Planning 132
5.2 Review of Variability of Reaction Times of Multimodal Gaze-Shifts 134
5.3 Review of 3D Kinematics of Head-Free Gaze-Shifts 135
5.4 Implications for a Complete, Multisensory Cognitive-Motor System 137
5.4.1 A single program of gaze-shift control, encoded in prefrontal cortex 137
5.4.2 Attentional control and realization of working memory 138
5.5 Implications for Neurophysiology of Multisensory Processing 139
5.5.1 Different levels of decision making realized in frontal cortex 139
5.5.2 Cortical and collicular connections of basal ganglia 140
5.6 Practical Implications 141
5.7 Concluding Remarks 142
5.7.1 Conclusions 142
5.7.2 Future Directions 142
6 References 144

1 General Introduction

We detect a limited range of signals from the environment. Basic sensory features and perceptions, e.g. for position, orientation, texture, color, form, and etc., are constructed in primary sensory cortex. At a higher hierarchical level, more complicated perceptions are realized in distributed networks by specifically associating sensory representations i.e. binding different features into a unique percept. This integration (binding) of sensory information may be controlled or commanded by various cortical units. It may be controlled by a posterior occipital / parietal network underlying a perception and be used for categorization and recognition. It may also be controlled by a frontal network underlying an action and be used for driving its executive parameters.

In this thesis, we consider how spatial information, received from visual and auditory sensory apparatus, may be integrated for perception and action. We specifically focus on the problem of making a gaze-shift to the inferred cause of the cross-modal stimuli. This problem is broken into three levels of information processing: 1) at the highest level, we are concerned with spatial perception of the cross-modal stimuli. Here we consider a task where a human subject is asked to report if the presented cross-modal stimuli belong to a same source or separate sources. 2) At a lower stage, we are concerned about when to implement the gaze-shift towards the inferred cause. Here we consider a task where the reaction times of gaze-shifts towards the perceived position of a target, based on cross-modal cues presented about the target, are recorded. 3) At the lowest level, we are concerned with how the gaze-shift is realized through a coordinated motion of eyes and head. We formalize the first level as a problem of dynamic inference, and propose a computational model of the cortico-striatal circuitry as its underlying representational structure. The second level is formalized as a computational model of cortico-collicular and striato-collicular connectivities. The third level is formalized as a computational model of brainstem oculomotor circuitry. We are not concerned with neurophysiological verifications in this thesis.

In this General Introduction, firstly, I review the experimental evidence about causal inference, reaction time variability, and the 3D eye-head coordination. Secondly, I review neurophysiology of cortical and subcortical brain networks underlying perception, action, attention, working memory reasoning, and decision making. Thirdly, I review the behavioral and neurophysiological evidence about eye-head coordination for gaze-shift. Fourthly, I review the computational architectures proposed about how the brain's neurobiology may give rise to cognition, and converge to Neural Engineering Framework as our selected approach. Finally, I frame the problems we address in this thesis as a hierarchy of information processing, starting at cognitive processing for causal inference at higher cortical areas, and converging to motor processing for eye-head coordination in subcortical areas.

1.1 Review of Multisensory Phenomena Addressed in this Thesis

1.1.1 Inference of the cause of the cross-modal stimuli

Sensory systems detect different types of signals originating from objects in the surrounding environment. For example, visual signals represent the electromagnetic waves with a specific range of frequencies that can be detected by the visual system, whereas auditory signals represent the mechanical waves with a certain range of frequencies that are detectable by the auditory system. These sensory received signals activate the corresponding sensory representations in the brain, which code for various features in space and time. Causal inference in animals is the process of estimating what events in outside world has caused the sensory representations in the brain (Shams and Beierholm, 2010, Lochmann and Deneve, 2011).

Here we consider a case where spatial position signals are available from visual and auditory senses. Causal inference reduces here to judging whether the two signal originate from one same object or two different objects. Experimental evidence shows that if the temporal features are very close and similar to each other, one may perceive an illusory common cause despite the mismatch between their spatial features (Vroomen et al., 2001a, b, Godfroy et al., 2003). Similarly, if the spatial features are very close and similar to each other, you may perceive an illusory common cause despite the mismatch between their temporal features (Vroomen and Keetels, 2006, Vroomen and Stekelenburg, 2011). These

spatial and temporal ventriloquism effects break down at large spatial or temporal disparities (Slutsky and Recanzone, 2001, Wallace et al., 2004).

In one study (Alais and Burr, 2004) observers were asked to report if a test stimulus, flash and click in spatial conflict, or a probe stimulus, flash and click presented together spatially, appeared more rightward. The main parameter they changed was the quality and spatial reliability of the visual signal. The subjects were told that the flash and click in the test stimulus belong to a unique object. So, the position where they perceive the unique object depends on the reliability of the signals that they assume they receive from it. For the case of sharpest visual stimuli they observed the classical ventriloquist effect such that the subjects perceive the object close to the position of the visual stimulus. For heavily blurred visual stimuli, they perceive the object close to the auditory stimulus. For intermediate levels of blurriness, they perceive the object somewhere between the positions of the visual and auditory stimuli. Their results imply that, when the subjects assume a common cause for the cross-modal stimuli, an intermediate position closer to the more reliable of the stimuli, is perceived as the location of the common cause.

Other studies let the subjects decide whether the two cross-modal signals belong to the same object or not (Slutsky and Recanzone, 2001, Wallace et al., 2004). They changed both the spatial and temporal relationships between the two presented targets. For very short-duration and synchronous stimuli, the subjects reported a unique object as the source of the signals and perceived it at the weighted average of the position of the two signals. By extending the presentation time or by introducing some temporal disparity between the signals the chance of reporting a unique cause for two spatially disparate signals decreased drastically. Also for synchronous stimuli with fixed and significant presentation time, increasing the spatial disparity between the stimuli decreases the percentage of the trials where the subjects reported a unique object as the source. Their results showed that when subjects are not told to assume a common source for the stimuli, they may localize the stimuli in common or separate spatial positions, and this decision is affected by the spatial and temporal features of the stimuli.

1.1.2 Variability of reaction times for gaze-shifts towards cross-modal stimuli

When we avoid reactionary motor responses towards sensory stimuli, we allow ourselves to plan actions based on a more complete set of information inferred from the sensory evidence. Such inference extends our perception beyond the sensory information, and provides us with a wider range of action synergies than sensory-driven reflexive movements. Selection of one of such action plans and the timing of its execution, then, depends on high-level cognitive processes, rather than reactionary sensorimotor paradigms. Such executive reaction time (RT) has been used to investigate hypotheses about the mental and motor processes to implement different tasks (Sternberg, 1969). In multisensory integration (MSI) research specifically, RT has been used to assess how combining multimodal stimuli with various intensities affect task implementation and response generation (Hershenson, 1962, Rubinstein, 1964).

It is well known that bimodal stimuli, e.g. visual and auditory, affect the reaction times of goal-directed saccadic eye movements. In particular, when the two stimuli are aligned in space and time, a considerable reduction of the saccade RT is typically observed relative to visual stimulus alone or to auditory stimulus alone. Conversely, RT increases more slowly or even decreases when the stimuli are presented farther from each other or when the delay between them gets larger (Frens et al., 1995, Corneil et al., 2002, Diederich and Colonius, 2004, Navarra et al., 2005, Diederich and Colonius, 2008a, b, Navarra et al., 2009, Van Wanrooij et al., 2010).

1.1.3 Eye-head coordination for shifting the line of sight

Gaze-shifts, i.e. rapid reorientations of the line of sight, are the primary motor mechanism for re-directing foveal vision and attention in humans and other primates (Bizzi et al., 1971b, Tomlinson and Bahra, 1986a, Tomlinson, 1990, Guitton, 1992, Corneil and Munoz, 1996). The fundamental aspects of gaze control kinematics can be addressed even in one dimension. They include the amplitudes and temporal sequencing of eye and head motion (Tomlinson and Bahra, 1986b, Guitton and Volle, 1987, Guitton, 1992, Sparks et al., 2002). The typical sequence of events includes a saccade, followed by a slower head movement and a compensatory VOR (vestibule-ocular reflex that keeps the gaze stable during some head movement by moving the eyes in the head). The aspects of this progression that we

will explore here include the variable timing of saccade, head movement and VOR, the influence of initial eye and head orientations, relative magnitudes of the contribution of these different phases to the gaze-shift and where the head falls in space after the gaze-shift.

Additional complexity emerges when one considers gaze-shifts from a two-dimensional (2-D) perspective. For example, the eye and head provide different relative contributions to horizontal and vertical gaze motion, which must be predictably accounted for saccades to produce accurate gaze shifts (Freedman and Sparks, 1997, Goossens and Van Opstal, 1997), and for the eye and head to end up in the right positions after the VOR (Crawford and Guitton, 1997b, Misslisch et al., 1998).

Finally, gaze control reaches its highest degree of complexity in 3-D (Glenn and Vilis, 1992, Crawford et al., 2003). First, there is an added dimension of motion control: torsion, which roughly corresponds to rotations of the eyes and/or head about an axis parallel to the line of sight pointing directly forward. Torsion influences direction perception for non-foveal targets (Klier and Crawford, 1998), binoclular correspondence for stereo vision (Misslisch et al., 2001, Schreiber et al., 2001), and must be stabilized for useful vision (Crawford and Vilis, 1991, Fetter et al., 1992, Angelaki and Dickman, 2003). More fundamentally, a 3-D description requires one to account for the non-commutative (order-dependent) properties of rotations (Tweed and Vilis, 1987, Hepp, 1994). These properties influence not only ocular torsion and the degrees of freedom problem, but also gaze accuracy, for reasons related to reference frame transformations.

1.2 Review of the Neurophysiology of Cortical Networks

The ultimate goal of neuroscience is to figure out how the brain categorizes information and represents those categories. All cognitive functions are physically realized through representations of knowledge in assemblies of neurons in the brain (McClelland and Rogers, 2003). Neurons are interconnected in various ways. Some connections are segregated in order to process information differently in parallel pathways. Some connections are convergent from many onto few nodes for integrating information. Some connections are local between all neurons of a network in order to retain and process some temporarily available sensory information (McLeod et al., 1998).

All cognitive functions can essentially be explained in the context of knowledge categorization (Hayek, 1952). Perception consist in classifying of objects by binding of features that have occurred together in time and space in the past. Action consists of associating a set of movements that, have previously learned, can attain some goal. Attention is allocating processing power on a specific class of perceptual or executive information. Memories are stored codes of how to instantiate different classes of perceptions or actions. Language is a sequential code (syntax) that can be decoded by the brain for creating the distributed network of some perception or action (semantics). Reasoning is modifying a class based on new information or based on its overlap with other classes, according to its learned logical rules of association (Fuster, 2005). Here we review the brain organization underlying these cognitive functions, leading to proposing a computational cognitive architecture in the next section.

1.2.1 Hierarchical organization of perceptual networks in posterior cortex

Perception is our representation of the world through our senses. The sensory foundation of perception is not a matter of controversy. However, some view perception to be reducible to effects of sensation through receptors and nerve cells (Boring, 1942). Such a viewpoint on perception, limiting it to sensory analysis of physical features (Locke and George Fabyan Collection (Library of Congress), 1690), however, is not complete as it ignores the subjective and historical aspects of perception (Helmholtz and Southall, 1924). All percepts are formed by classifying the sensory information according to past memories (Berkeley, 1709). Even sensations can be thought of as retrieving ancestral or phyletic memory, genetically coded and inherited within species. Perceiving is, then, remembering, updating of memory. This active nature of perception has been in large part dismissed by psychophysicists, because of their emphasis on sensations, and by cognitive scientists, because of their emphasis on symbolic essence of cognition.

The other aspect of perceptual processing ignored by psychophysicists and cognitive scientists is the parallel and unconscious execution of its information processing. Vast majority of testing and verifying hypotheses, underlying categorization of sensory information for perception, takes place along multiple channels and outside of consciousness (Barsalou, 1999). This matching act of sensory impressions to pre-

established memories, is mostly aided by attention. Guided by memory or pre-conceived search plans, attention determines the course of categorization, by its two main functions: an inclusion component that allocates the limited processing power to updating some relevant perceptual memory; and an exclusionary component that attenuates processing in other irrelevant sectors. In case of unsatisfactory matching, the selected perceptual memory is modified, or another percept is projected on the present reality.

The central question, then, is how one perception is constructed, is segregated from others, and preserves its identity despite discontinuities. Focusing on visual spatial perception, Gestalt psychologists developed certain principles of such organization, for example, based on proximity, similarity, continuity, and closure (Koffka, 1935). These principles explain grouping of elements in a gestalt based on regularities in the spatial relations between the elements. Such regularities reflect on existence of a more abstract code for a whole, which is bigger than, and independent of, the sum of the parts (Anderson, 1995). This idea can be applied to temporal domain, other sensory modalities, and multiple levels of abstraction. Then, in general, a perception can be defined as a specific relational regularity, in time and space, which is discovered between elementary parts. This, at least at a phenomenal level, explains the problem of perceptual constancy. What defines a perception is a special relationship between its elements, not the detailed features of each element, e.g. sizes and directions.

The underlying neurophysiology of perception is the activation of a hierarchically organized network of connected neural populations, distributed all through the posterior cortex. The distributed network represents a relational regularity in its associative structure (Edelman, 1989). At the lowest stage of the perceptual hierarchy in the posterior cortex, there exist primary sensory areas. These most peripheral areas are specialized to recognize attributes defined by the physical parameters of the objects, giving rise to sensory features of the world. The perceptual representations, at this level, are primitive and faithful to the environment (Gilbert, 1998). Networks in these unimodal sensory areas categorize simple percepts, within that modality, such as color, orientation, motion, pressure, and pitch. Individual features of newly presented sensory information are analyzed, separately and in

parallel, to form structures in time and space. These analyses are then sent to higher unimodal association cortices.

In higher areas of unisensory association, the separately analyzed features, recognized from environmental phenomena, are integrated to form more complex features, within the corresponding sensory modality (Felleman and Van Essen, 1991). Neurons in these areas have broader receptive fields and are responsive to more complex patterns and forms in the environment. Such cognitive networks represent universal and consistent relational regularities, in time and space, to categorize the simpler features, formed in primary sensory areas, into more complex and abstract classes. Such complex classes may have associations to other areas, at the same or different levels, to identify that percept more personally and specifically. For example, face recognition areas identify a face by classifying spatial features in a very specific way. Then each instantiation of the class "face" has specific emotional connotations, realized through associations with limbic cortex and amygdala.

The results of unisensory analyses in primary and associative cortices converge onto large trans-modal posterior cortex (Mesulam, 1998). Lesions of this cortex leads to agnosia or semantic aphasia. Cognitive networks in these areas represent the perception of highly abstract symbols. Symbols abstract the essential features of a percept across wide variations of its sensory instances. A cat, a desk, an opera, or a rap song, all can be identified in the real world in a wide range of variety. Symbols are amodal in the sense that, in their relational structure, they may include associations with lower-level percepts from multiple sensory modalities; and they may be instantiated by recognition of lower sensory features, from environment, within any of those modalities (Barsalou, 1999). Symbols, obviously, have non-sensory dimensions, e.g. affective and emotional, realized through their associations with such categorizations in limbic system.

Multiple patterns of connectivity, within and between hierarchical levels, serve as different ways to represent and process percepts in distributed networks (Edelman et al., 1978). Upward convergence from multiple lower-level areas to a single higher-level area has been observed, which may underlie binding of different features into a higher percept. Upward divergent connections from a single lower-level population to multiple higher-level areas

has been observed, which is thought to be involved in associating a common property to different categories of perception. Collateral connections between areas of the same hierarchical level has been observed, which underlies binding of percepts at the same level. Local recurrent connectivity in neurons of one population is ubiquitous in cortex and underlies retaining and temporal processing of information within that population. Global recurrent connectivity has been observed from higher to lower areas, which serves the attentional top-down processing.

1.2.2 Hierarchical organization of executive networks in frontal cortex

Translation of perception to action is mediated by projections from sensory structures to motor structures, at all levels of the two hierarchies, across the central nervous system (Young, 1993). The most reflexive reactions take place through interneurons between sensory and motor nerves in the spinal cord. Goal-directed actions, though, depend on cortical connections. Routine and automatic behaviors are realized in transformations between primary sensory and motor cortices. More complex motor actions involve the higher-order, associative, cortical areas. The complex motor synergies consist of temporal organization of percepts and actions. This integration of lower-level, automatic actions, along with updating of the associated percepts, is the cognitive function of the prefrontal cortex, at the summit of the executive hierarchy in the frontal lobe. This role is fulfilled through massive reciprocal connections of the posterior cortex and lower-level frontal areas with the prefrontal cortex (Fuster, 1997).

The motor cortex (M1) is at the lowest level of the cortical executive hierarchy in the frontal lobe. Neural populations in M1 encode directions for the movements of specific body parts (Iriki et al., 1989). Each sub-nucleus in M1 controls a group of muscles that can move a body part in any direction (Georgopoulos et al., 1982). The specific firing pattern of each sub-nucleus, then, leads to a specific activation pattern in the corresponding group of muscles, which moves the body part in a specific direction (Sato and Tanji, 1989). Therefore, M1 representations are not organized somatotopically, but rather based on the intended action.

In a higher level of the executive hierarchy, in premotor cortex, more global aspects of movement are represented. On one hand, in premotor cortex proper (area 6b), cells are

attuned to the kinematic properties of movement, especially trajectory (Weinrich and Wise, 1982). To compute such variables, this nucleus receives connections from posterior areas involved in representation of space (Crutcher and Alexander, 1990). On the other hand, the so-called mirror units in premotor cortex, seem to encode movements of other subjects in the environment (di Pellegrino et al., 1992). To calculate these patterns, this nucleus is in constant talk with posterior areas involved in motion perception. Finally, in upper and more medial parts of premotor cortex (area 6b), the so-called supplementary motor area, cells are activated when a sequence of moments is executed (Mushiake et al., 1990). Neurons are tuned to the sequence rather than a specific movement component. This indicates a higher level of executive abstraction within time dimension. Temporal coordinates of sequence is encoded here, rather than the spatial coordinates of the movement.

At the highest level of the executive hierarchy, the prefrontal cortex is involved in representing complex programs of action (Quintana and Fuster, 1999). Such programs consist of integration, across time, of multiple actions with perceptual referencing and updating, in order to achieve some abstract goal (Petrides et al., 2012). Lesioning of prefrontal cortex causes deficits in learning to formalize action plans, by temporal integration of sensory and motor information. Such lesions often also lead to deficits in syntactical formation of linguistic sequences. Neurons in prefrontal cortex show sustained activity between cue onset and target onset in memory-delay tasks (Quintana et al., 1988, Yajeya et al., 1988). This firing patterns has been assigned to working memory, the attentive process of temporal integration of information to plan an action. Another observation is that, while prefrontal cortex is active during learning sequential movements, the activation disappears when the action becomes automatic and routine, and activity of basal ganglia increases. Well-established action programs seem to be relegated to lower structures.

The general abstract expertise assigned to prefrontal cortex is formulating a program of actions and perceptions to serve a goal (Fuster, 2005). But, what is an example of such complex programs, formalized by integration of multiple actions and selections through time? The reaction to appearance of a novel, salient, possibly dangerous stimulus in the peripheral visual field is one such program. The first element in this program is to use the

retinal image for reorienting the line of sight towards the stimulus. To do that, a massive subcortical network coordinates movements of the head and the eyes, to change the gaze orientation (reviewed in the section 1.3). The second element of the program is to recognize the stimulus based on the visual information available through the new pattern of retinal stimulation. This is achieved in posterior parietal cortex, within the massive networks of the visual perceptual memories. The chosen perceptual network is kept active, by attentional control in working memory, for perceptual updating and recognizing an appropriate action. The third element of the program is to select an action based on recognition of the stimulus. This could be running away, looking away, smiling, punching, etc. based on the nature of the stimulus.

1.2.3 Attention

Attention, in psychology, is characterized as the focusing on one out of many possible, perceptual or executive, information processing paradigms (James, 1890). Attentional control does not have to be conscious, as much like the majority of neural information processing in the brain. The brain has, as the essence of its cognitive functionality, an enormous number of overlapping, intersecting, and interacting networks of neural populations, underlying a vast array of complex patterns of perception and action. Allocation of neural resources to one of such networks, at the expense of withdrawing others (Uexküll, 1926), is the core of the cortical processing we call attention.

Attention is an essential and inherent component of any cortical neural processing (Neisser, 1976). Every associative network of neural populations, representing and processing sensory or motor information, has as its core, the capabilities and connections to both exert attentional control and accept it. There is no evidence of a separate structure in the brain dedicated to attention as an independent function. Attention is the selective activation / deactivation of perceptual and motor networks, in a timely fashion, by some other strongly activated network, to serve the purpose of that network. When a program of action-perception sequence is formulated in prefrontal cortex, it is also learned when and to what networks to send attentional signals. When such programs become routine and automatic, they are relegated to lower-level areas, along with their attentional commands. So, attentional signals need not be top-down. For the complex programs that need adaptation

during the task, attentional signals come from the top, namely prefrontal cortex. For rapid automatic programs, the attentional control originates from lower structures.

As clear from its definition, attention has two faces; an inclusionary role and an exclusionary role. This has been shown to be realized by selective excitation and inhibition of the target nuclei, at all levels of the central nervous system (Kuffler and Nicholls, 1976). At the lower motor levels, in spinal cord, reciprocal innervation is one such mechanism. Imagine one motoneuron that innervates a flexor muscle, and another one that innervates an extensor muscle. They are parts of a network that controls the movement of the leg at the knee. This network also includes a sensory afferent from the spindles of the extensor muscle to the flexor motoneuron, and another sensory afferent from the spindles of the flexor muscle to the extensor motoneuron. When the leg is moved up, the extensor motoneuron innervates the extensor muscle, and extensor sensory neuron inhibits the flexor motoneuron, all as part of one action. At the lowest sensory level, in the lateral geniculate body, the on-center-off-surround receptive fields embody another such attentional mechanism. Such LGN neurons fire harshly if light is shown at the center of their receptive field, but show absolutely no firing if an annulus of light is shown around a dark center of the receptive field.

Perception is matching of sensory information to associative networks of perceptual memory (Grossberg, 1999). Perceiving is a program devised by integration of multiple elements across time, which has been relegated, as an automatic process, to lower sensory areas (Moran and Desimone, 1985). The first element of such program is to identify primitive perceptual features like color, orientation, position, pressure, pitch, etc. Attentional control in this element is to excite one instance of such features in favor of inhibiting others. If it is green, it is not red or blue. If it is at the top-right corner, it is not at the bottom or left. Next element is to identify more abstract forms based on the primitive ones. One attentional control here is the selective excitation of the best choices at the expense of inhibiting others. It is a face, so, it is not a chair. It is a melody, so, it is not spoken language. Another, this time top-down attentional control is from the more abstract forms on the primitive features. If it is a face, let’s not think the line orientations are very straight. If it is music, let’s not notice the large frequency differences in successive pitches.

The next element is to recognize the stimulus symbolically, if it is possible, based on all the features, primitive or abstract, formed at lower levels. The heterarchical attentional control is to instantiate the associative networks of one symbol in favor of other possible ones. If this is John's face, it is not Jack's. If this is a Led Zeppelin song, it is not a Tool song. The top-down attentional control, here, originates from the high-level symbol on the lower-level features. If it is Led Zeppelin, let's adapt the parameters of the lower-level perceived features based on our memory of Led Zeppelin's style of music.

1.2.4 Working memory

We mentioned how a program of action-perception is formulated on the time axis, in the prefrontal cortex (see section 1.1.2). We also mentioned that such programs, when become routine and automated, are relegated to lower-level sensory or motor areas. Top-down or bottom-up, automated or not, when a program is executed through time, various elements of the program send attentional control signals to specific neural populations, to retain their represented signal, or process it in some special manner, during the time of the execution of the program (Fuster, 2005). What we call working memory refers to this attentional control process. The essential properties of working memory are those of a perceptual or executive memory, which is held active, in the focus of attention, as required by information processing underlying the prospective action.

In tasks with memory delays, some neurons, found all over the brain depending on the task, show sustained strong firing during the delay interval (Fuster and Alexander, 1971). This delay activity was strictly dependent on the task requirement to act contingent on the memorized signal. It was also not induced just by expectation of the reward. Such characteristics made these cells likely candidates as being engaged in the process of attending to a perceptual / executive memory, which forces it to retain or process the recent instance of itself (i.e. working memory). Their sustained activity is vulnerable to distraction, and its level is correlated with efficacy of the stimulus. All such cells belong to an associative network that represents the perception of the memorized signal. Maximum of such persistent activities belong to the part of the network that represents the features of the percept that change trial by trial, and indicate the next course of action.

During execution of a memory task, a large distribution of neurons get activated, in frontal lobe because of the execution aspects, and in the posterior cortex due to the perceptual aspects of the task. In the course of the task, maximum sustained firing moves between regions according to the task's immediate demands. A reasonable interpretation is that all these neural populations are parts of a single distributed network that represents the program of the task with its motor and sensory components. Working memory is the joint activation of different parts of this network during performing the task. Prefrontal cortex is active all through each trial because it codes the temporal relations of different stages of the task. Posterior areas are active in some stages due to the need for perceptual processing in those stages. Lower frontal areas are active in some stages because of the need to execute some actions in those stages. The studies of simultaneous recording from lateral prefrontal cortex or the inferotemporal cortex, and cooling of the other, support these interpretations, among others (Fuster et al., 1985).

1.2.5 Reasoning and Inference

We reviewed how the perceptual and executive processing are organized in the brain. Can we extend this framework to how deductive reasoning finds the unknown (not available through senses) features of a concept, and how inductive reasoning creates new forms of perception and action. The abstract structure of a perceptual or executive memory is at the core of any intelligent functions that are defined for it. This structure determines what lower-level components it can accept, and what higher-level networks it can be part of. In essence, this structure is the logic of associations for the corresponding network.

Let's first consider deductive reasoning. It is the process of logical reasoning to find unknowns about a percept or action based on the structure of its associations (Johnson-Laird, 1999). It is this associative logic of a percept or action that identifies the questions that could possibly be asked about it. This logic characterizes the unknown parts of the percept or action, and how such unknowns may be inferred from associations with other overlapping memory networks instantiated by the sensory signals, or by planning an action which will help us gather the appropriate information. For example, we see the front side of a car. We recognize it as a car just by this very limited visual information. We never ask how its wings look like because, from the structure of our perception of a car, we know it

does not fly. But we may ask what kind of engine it has. We may try to answer this by referring to our memory of the engine type of these specific cars. Or we may plan an action to go and look at the engine. In any event, such reasoning depends essentially on the internal structure of the network.

Another form of intellectual behavior is problem solving by inductive reasoning. It is the process by which we conclude that what is true for some observational instances is true for a more abstract class (Sternberg, 1985). Within our reviewed framework, inductive reasoning consists of creating a new higher-order symbol constructed of a network of associations of features and behaviors and logic, all abstracted based on multiple repetitive observations. Imagine we want to answer the question of how kangaroos move around. We go observe a number of kangaroos moving. We identify the similarities in our observations. We match the shared characteristics with a pre-existing perceptual memory, namely jumping in a specific way. We abstract our observations by adding this general way of jumping to the distributed network of our memory of kangaroos. Analytical reasoning and finding similarities, besides the abstraction through matching the similarities to a higher network, constitute the core of inductive reasoning (Holyoak and Thagard, 1996).

Such intellectual behaviors of problem solving by deductive and inductive reasoning, although dependent in their specifics on the context and nature of the question, can be thought of as independent and abstract functionalities. They are essentially executive programs of integration of perceptions and actions on the time axis. They include selective, sequential allocation of different percepts or actions to different stages of a program with the goal of gathering the right information to answer a specific question. This implies the likely role of executive networks in the prefrontal cortex in intelligent behavior (Duncan et al., 1995, Duncan et al., 1996). Such networks contain vast connections to lower levels of executive processing in frontal lobe structures, and to all levels of perceptual processing in posterior cortex. It has been experimentally shown that the most consistent correlate of cognitive abilities, in an intelligence test, is the coherence in the frequencies in the theta range, recorded from frontal and posterior cortices (Anokhin et al., 1999). Intellectual capacity, therefore, has been associated to synchronicity in wide cortical areas, which

implies recruitment of a large number of cognitive faculties in the service of a program, whose goal is to answer a question.

1.2.6 Decision making

Complex programs, in prefrontal cortex or relegated to lower cortices, consist of calculated succession of perceptual and executive processing. Such programs sometimes include multiple possible courses of action in a next stage, selected based on gathering some sensory information at the current stage. This is the essence of decision making in the brain. Decision making is ubiquitous at all levels of cortical neural processing. It is part of lower-level networks, in sensory or motor hierarchy, if its corresponding program is routine and automatic. It is controlled by the prefrontal cortex if its program is adapted to solve a novel situation. However, decision making, in its essence, consists of general components: 1) what set of alternatives are considered possible, in general, to govern the next course of action, 2) which subset of possible alternatives are considered viable in a specific situation, 3) how a decision is made between multiple viable alternatives, 4) when the selected course of action is initiated.

In any event, the decision to act in a certain way depends on perceptual processing of sensory information, and posterior connectivities are one of main players. Cells in middle temporal (MT) area of posterior cortex respond selectively to moving visual stimuli (Britten et al., 1996), in a decision-making task. Similar MT neural behavior has been observed when the direction of ambiguous moving gratings has to be recognized (Logothetis and Schall, 1989). Also, cells in somatosensory cortex show discriminatory firing in response to different frequencies of mechanical vibrations (to fingertips), when trained to report differences in such perceived vibration frequencies (Hernandez et al., 2000). Perceptual processing in the posterior cortex, therefore, seems to be a cornerstone of decision-making. Such posterior areas include the whole decision structure within themselves, if the decision-making is part of an automatic process. They send their results to executive networks of prefrontal cortex, if the decision-making is part of a program to solve a novel problem.

Frontal executive networks constitute the second component of decision-making processes (Damasio, 1994). Every part of the frontal cortex is embedded in a large amount of

reciprocal projections, both excitatory and inhibitory, with posterior cortex and subcortical areas. This supports the ubiquity of decision-making and weakens the idea of convergence in some central unit (Fuster, 1997). Therefore, decision making, included in many frontal networks, is governed by multiple influences arriving from various cortical or subcortical sectors. Results of decision is applied through selective inhibition, controlled by the executive network, of structures that represent the alternatives for the next course of action.

1.3 Head-Free Gaze-Shifts in Three-Dimensional Space

1.3.1 Kinematics of eye-head coordination

Gaze-shift is a rapid reorientation of the line of sight to redirect attention and vision (Bizzi et al., 1971, Tomlinson and Bahra, 1986a, Tomlinson, 1990). Natural gaze-shifts are implemented by complex coordination of eye and head movements, in time and space. The three motor mechanisms, which are coordinated to realize a gaze-shift, include an eye-in-head saccade, a more sluggish head movement and a vestibule-ocular reflex (VOR) (Tomlinson and Bahra, 1986b, Guitton et al., 1990, Freedman and Sparks, 1997, Roy and Cullen, 1998). A kinematic strategy for this coordination is the subject of the third part of the thesis (chapter 4). Here we review the some experimental findings about such strategy.

One aspect of this strategy is in the temporal and sequential domain. The typical sequence of events, experimentally observed, includes a saccade, followed by a slower head movement and a compensatory VOR eye movement. The timing of the saccade, head movement and VOR may be variable. The magnitudes of the movements influence the timing. Also, the initial orientations of eyes and head affect the timings observed.

Another aspect of the coordination strategy is the amplitude of the different motor movements. Eyes and head have been shown to provide different relative contributions to horizontal and vertical components of gaze-shift. Also the amount of VOR eye movement is variable based on the initial eye and head orientations. All these variabilities must be predictably accounted for saccades to produce accurate gaze shifts (Freedman and Sparks, 1997, Goossens and Van Opstal, 1997), and for the eye and head to end up in the right positions after the VOR (Crawford and Guitton, 1997a, Misslisch et al., 1998).

The most important aspect of the coordination strategy is in the spatial domain, and how the degrees-of-freedom problem is dealt with. An infinite number of rotational axes can be employed to bring a rigid-body from any given initial orientation toward a final 2-D direction, each resulting in a different amount of final torsion. Donders' law states that only one final eye orientation is achieved for each 2-D gaze direction, and thus only one axis of rotation can be used (Glenn and Vilis, 1992, Crawford et al., 2003a). Orientation of the eye relative to the head and orientation of the head relative to the shoulder obey Donders' law at their stable orientations, when the head and body are in normal upright postures (Misslisch et al., 1994, Klier and Crawford, 2003). Orientation of eye-in-head has also been shown to obey the Listing's law (Ferman et al., 1987b, a, Tweed and Vilis, 1990, Straumann et al., 1991); If torsion is defined as rotation about the axis parallel to gaze at the primary eye position, then Listing's law states that eye orientation always falls within a 2-D horizontal-vertical range with zero torsion known as Listing's plane (LP). Note that in order to maintain eye orientation in LP, rotations must occur about axes that tilt out of LP as a function of eye position, a phenomenon known as the half angle rule (Tweed and Vilis, 1990). In contrast, orientation of head-on-shoulder has been shown to obey the Fick strategy (Glenn and Vilis, 1992, Crawford et al., 1999, Klier et al., 2007) where horizontal rotation occurs about a body-fixed vertical axis, vertical rotation occurs about a head-fixed horizontal axis, and the remaining torsional component is held near zero. Mechanical factors appear to aid these constraints by implementing some of the position-dependencies required to deal with non-commutativity. In particular, eye muscles appear to implement the half-angle rule (Demer et al., 2000, Ghasia and Angelaki, 2005, Klier et al., 2006). However, it is clear that mechanical factors do not constrain eye and head torsion, because the eye violates Listing's law during the VOR (Misslisch et al., 1998, Crawford et al., 1999, Glasauer, 2007) , and the head constraint can be violated voluntarily or when used as the primary mover for gaze (Ceylan et al., 2000).

1.3.2 Subcortical networks of eye-head coordination

In response to a signal that codes the position of a spatial target in eye-centered coordinates, originating from the superior colliculus or the frontal eye field, the brainstem circuitry plans and generates coordinated movements of eyes and head, in order to change the line of sight towards the target (Sparks, 2002). These brainstem nuclei, accordingly, communicate a

pattern of activity to motoneurons that innervate eye and neck muscles, for them to move the eyes and the head in a specific way. The mechanisms, representations and connectivities underlying head movement control and its coordination with saccade and VOR in 3D space are largely unknown (Chen and Tehovnik, 2007). Here we review the little we know about the neural circuitry of saccade generation in brainstem.

The eyes rotate by the action of three pairs of muscles. Horizontal rotations are generated by medial and lateral rectus muscles. Vertical rotations are produced by superior and inferior rectus and superior and inferior oblique muscle pairs and torsional movements are produced by contractions of combinations of superior/ inferior rectus and superior/inferior oblique muscle pairs (Robinson, 1964, 1973, 1978). Different motor neurons innervate each muscle (Fuchs and Luschei, 1971).

Ocular motoneurons have a burst-tonic discharge pattern, tonic coding for eye orientation, burst coding the saccade (Robinson, 1964). This pattern is said to be resulted from their connectivities with saccade burst generators (SBGs) in reticular formation (RF). The tonic activity is proportional to the eye orientation. This tonic activity originates from the tonic neurons of the neural integrators in interaction with vestibular nuclei (Klier et al., 2002). These neural integrators are located in the interstitial nucleus of Cajal (INC) providing the vertical eye position signal, and in the nucleus prepositus hypoglossi (NPH) providing the horizontal eye position signal (Sparks, 2002).

The burst activity of the oculomotor motoneurons is proportional to the saccade amplitude. These bursts of activity originate from two distinct SBGs, which are composed of neuron types with specific activity patterns. The first class includes medium-lead burst neurons (MLBs), which emit bursts of discharge before saccade onset, during ipsilaterally directed saccades. There are two subtypes of MLBs (Sparks and Travis, 1971, Cohen and Komatsuzaki, 1972): 1) Excitatory burst neurons (EBNs), which code for eye velocity and acceleration. They project to ipsilateral motoneurons and the neural integrators. 2) Inhibitory burst neurons (IBNs), which inhibit the contralateral motoneurons, and also project to tonic neurons of the neural integrators. The inputs to both EBNs and IBNs originate from long-lead burst neurons (LLBs) in both SC and brainstem. LLBs emit bursts

before saccade and reach their maximum firing at saccade onset (Scudder et al., 2002). They are selective to the direction and amplitude of the saccades.

Burst neurons in the rostral interstitial nucleus of medial longitudinal fasciculus (riMLF) also provide monosynaptic excitatory input to the motor neurons that are involved in torsional rotations of the eye. The right and left riMLFs both contain burst neurons with up and down direction selectivity, but the right riMLF has preference for clockwise movements whereas left riMLF has preference for counter clockwise movements. Therefore, torsional eye movements can be generated by a balance of activity between up and down neurons with the same rotational preference (Hepp et al., 1988, Crawford et al., 1991, Henn et al., 1991).

Although the commands for horizontal and vertical components are generated in different regions of the brainstem (and are seemingly independent), during oblique saccades with both vertical and horizontal components, the onsets and the duration of the two components are synchronized (Guitton and Mandl, 1980). They project to and inhibit MLBs, functioning as saccade temporal switch. They discharge tonically during fixation and stop during the saccade (Pare et al., 1994).

1.4 Review of Computational Neurocognitive Architectures

How come animals' cognitive systems are so incredibly robust, intelligent, and adaptive, such that it looks so far from the reach of human engineering? Is it because the cognitive processing is only realized completely and robustly by the biological mechanisms in the brain, and that we have a very limited understanding of the brain? Or does it seem so far from the reach of our knowledge because we are ignorant of the underlying complex mechanisms of animal cognitive structure? Or probably both, i.e. our inability to create a real cognitive system should be related to both our lack of understanding of the brain as a processing unit, and our shortage of knowledge about cognition as the abstract formal system?

Despite the long way for engineering to get even close to animal natural intelligence, for any reason, there has been a number of invaluable attempts to propose biologically plausible cognitive architectures. Here, in this section, we first review general defining

criteria for any computational framework to meet, to be considered a viable cognitive architecture. We then review a number of approaches to the problem of computational modelling of cognition. Next, we evaluate different approaches based on the general criteria and converge on a unifying framework. Finally, we summarize the basic features of the Neural Engineering Framework, our selected approach.

1.4.1 Criteria for evaluation of different architectures

The first criterion is that any representational capacity ascribed to a cognitive agent must be able to explain the systematicity of our thoughts. This refers to intimate link between some sets of representations (Fodor and Pylyshyn, 1988). Why is it that, if we can assign a property to an object, the same property can be assigned to other objects as well? Or why is it that multiple instances of a property can be thought to be assigned to a same object?

The next criterion is that any cognitive system should account for the meaning of a complex, represented concept, based on the meaning of its constituent representations, according to some logic of compositionality. This could be the associative logic of binding features into a symbol, or the syntactical logic of composing actions and perceptions (verbs and words in linguistics) into a complex program (sentence in linguistics). The better a computational cognitive architecture defines systems that meet compositionality criteria, the more robust it is in processing of novel, complex representations (Fodor and Pylyshyn, 1988).

Another criterion is the ability of a system to generate a vast number of complex regularities based on a few simpler ones, and rules to combine them. This is directly dependent on availability of representations of generic roles, abstracted from experience. In linguistic terms, this is the problem of grammatical templates, where a variety of words can play the same role, leading to numerous combinations. The only limit to this seemingly infinite productivity is limitations of time and memory resources (Jackendoff, 2002).

Another general ability of a cognitive architecture is to bind basic representations to form complex representations, and then, to be able to distinguish multiple instances of one complex representation as belonging to a shared structure (Jackendoff, 2002). A proper cognitive agent can construct appropriately complex structures, within the limits of time

and processing resources, and then use this generic structure to recognize its instances fast and robustly.

The problem just mentioned about binding and recognition for complex concepts, is also applicable to complex, time-integrative programs (or sentences in language). An example of this is how people are flexible in using the structure of language independent of its content (van Gelder, 1998). This refers to the availability of generic structures to integrate concepts on the time axis, and then being able to use that to distinguish a shared structure, leading to meaning, in numerous instances of it.

Human cognition is robust, i.e. it is not sensitive to damaged, missing and noisy data (Rumelhart et al., 1987). Our engineered computers are robust in one respect, that they use components (transistors) whose states are interpreted as 'on' or 'off', and use a large margin (±5 V) to separate those states (leading to massive use of energy). In other respects computers are not robust at all. Any novel situation in the environment is not accounted for by the hardware, i.e. the hardware cannot run a program that is written poorly. It cannot use past experiences to interpret a noisy input reasonably. However, natural cognition is robust in all those respects, without using large amounts of power, like computers do.

Cognitive systems are adaptable as they can use their memory appropriately. They can use their memories to recognize novel stimuli (Schoner and Thelen, 2006). They can adapt their memory based on new information. Any computational architecture should account for this fundamental capability. Different architectures account for adaptability in very different ways. The better approach is the one that realizes adaptability within the limits of the system on time and processing resources.

Any cognitive architecture needs to explain how to incorporate memory of past experiences into cognitive functions. Such ability is classically approached through introduction of long-term memory with static nature and large capacity, and working memory with dynamic nature and limited capacity (Cowan, 2001). In any event, cognition necessitates manipulation of complex, compositional structures within and using memory, and a cognitive framework needs to account for it.

1.4.2 Symbolic, connectionist, and dynamicist approaches

Symbolic approach, realized in production systems, is the classic viewpoint to modelling cognition. They usually consist of two core parts: 1) a set of if-then rules, 2) a control structure that matches the input to the "if" part of the rules to identify the next course of action by the "then" part. This approach was first used in the "General Problem Solver" (Newell and Simon, 1972), and then became the core of a number of attempts like Soar (Newell, 1992), Executive-Process/Interactive Control (Meyer and Kieras, 1997), and Adaptive Control of Thought (ACT-R) (Anderson, 1983). ACT-R is loyal to the basic structure of production systems, however, it assigns "utilities" to production rules to identify their likelihood of being applied. These utilities change over time with adaptation to the context and environment (Anderson, 1990, 2007). LISA (Learning and Inference with Schemas and Analogies) materializes the idea of synchronization, of spiking activity across a vast network of neural populations (von der Malsburg, 1995), as a method for representing structural relations underlying feature binding (Hummel and Holyoak, 1997, 2003). To avoid LISA's scaling problem, NBA (neural blackboard architecture) introduced a central, flexible, temporary structure underwriting symbolic binding, that can be accessed by many independent processes (van der Velde and de Kamps, 2006). ICS (integrated connectionist symbolic) architecture, on the other hand, uses Harmonic Grammar to implement optimality theory, a symbolic structure, in a connectionist network. It tries to model linguistic processing by tensor products (Smolensky and Legendre, 2006).

If you want to build a fast cognitive system–one that directly interacts with the physics of the world–then the most salient constraints on your system are the physical dynamics of action and perception. Roboticists, as a result, seldom use production systems to control the low-level behavior of their robots. Rather, they carefully characterize the dynamics of their robot, attempt to understand how to control such a system when it interacts with the difficult-to-predict dynamics of the world, and look to perception to provide guidance for that control. If-then rules are seldom used. Differential equations, statistics, and signal processing are the methods of choice for this "dynamicist" approach (van Gelder, 1998). According to this approach, cognitive systems can only be properly understood by characterizing their state changes through time. These state changes are a function of the tightly coupled component interactions and their continuous, mutually influential

connections to the environment. Dynamic constraints are clearly imposed by the environment on our behavior. The nature of that environment can have significant impact on what cognitive behaviors are realized (Healy and Rowe, 2014). Dynamic field theory, a complete instance of this approach, focuses on various types of stable attractor networks, and how they can be formed, to represent metric dimension (luminance, position, etc.). Networks of such stable patterns are created, by connecting them in various configurations, for modeling sophisticated behaviors, such as ocular control, reach preservation, and infant habituation.

The third approach to modeling cognition, i.e. connectionism, is based on connections of large number of very basic computational units in various patterns. Neurons, here, are reduced to nodes of such networks that perform a simple transformation of input to output. However, when sufficiently large number of such nodes are grouped together, their network function is interpreted as implementing rules, classifying patterns, and performing cognitively-driven behaviors (Hummel and Holyoak, 2003, van der Velde and de Kamps, 2006). As the computational and representational properties of such nodes bear small resemblance to real neurons, connectionist models cannot be directly compared to most data recorded from the brain. As an example of this approach, Leabra (local, error-driven and associative, biologically realistic algorithm) is a method for learning central elements of a cognitive architecture, by applying a k-Winner-Takes-All algorithm. It has been used to construct Primary Value and Learned Value model of learning (O'Reilly et al., 2007), which simulates behavioral and neural data on Pavlovian conditioning and the midbrain dopaminergic neurons that fire in proportion to unexpected rewards.

1.4.3 Reconciliation of different approaches

Cognition consists of sub-personal processes, such as effortful reasoning, that generate representations of the world that go beyond the information available to our senses (Eliasmith, 2013). This extends our perception and provides us with action synergies beyond sensory-driven reactions. Meanwhile, dynamic constraints are clearly imposed on our behavior by the environment. Cognitive systems can only be properly understood in case their state evolution through time is characterized (Healy and Rowe, 2014).

Classically, cognition is described as chains of if-then rules which statically transform the internal states of the system (Newell and Simon, 1972, Anderson, 1983). At the same time, classical control theory focuses on goal-directed motor planning within the time constraints of environmental interactions (van Gelder, 1998). While the former approach ignores the short-term dynamics of perception and action, the latter ignores the internal system, and sacrifices the high-level linguistic processes, such as complex planning and deductive reasoning. However, It is clear that the "dynamic perception / action" and the "high-level inference / language" are integrated in humans: cognitive animals. Interestingly, the way to solve this apparent conflict is, probably, through understanding how the brain encodes and transforms information in vast networks of neurons; the phenomena that traditional connectionism claims to understand, but not really.

Here, we adopt a unified approach where a model is identified by functions of both the internal state variables and the time. Inspired by the brain, such more general models are realized through a distributed network of parallel processing units. This approach simultaneously accounts for syntactic manipulations of representations underlying inference, and flexible control of information routing between different units through time (Eliasmith, 2013). Although we do not deal here with the neural implementation of the model, all the representations and transformations are designed based on the known neurophysiology, and can be neurally realized by a recent theoretical approach, neural engineering framework, which unifies the symbolic, connectionist, and dynamicist viewpoints (Eliasmith and Anderson, 2003, Eliasmith et al., 2012). The relatively high number of variables in such models is because we are modelling an adaptive, robust biological system which can behave and survive in an uncertain, changing environment.

1.4.4 Neural Engineering Framework

Cognitive, perceptual, and motor abilities have been associated to the activity of neural circuits in the brain. This mere association has convinced some that cognition will emerge if we can model single neurons' performance and connect them according to the real synaptic statistics (Markram, 2006). However, cognition has not yet emerged from data-driven large scale models, and there are good reasons to think that cognition may never emerge (Eliasmith and Trujillo, 2014). The Neural Engineering Framework (NEF)

(Eliasmith and Anderson, 2003) is an approach which starts modeling cognition by describing the system at a higher level of abstraction and then realizing it using neural models with adjustable degrees of accuracy. This method has its roots in the work of a number of researchers (Georgopoulos et al., 1986, Salinas and Abbott, 1994, Rieke, 1997). The models designed within this framework closely resembles physiological findings in activity of single neurons, timing of responses, and behavioral errors without being built into the model (Rasmussen and Eliasmith, 2013, Stewart and Eliasmith, 2013).

Within NEF, a group of neurons creates a distributed representation of a vector $\mathbf{x}(\mathrm{t})$ which is varying in time. A preferred direction vector $\widetilde{\mathbf{\Phi}}$, a bias current J_i^{bias}, and a scaling factor α are associated with each neuron i. If the nonlinearity of the neural model is $G[\cdot]$ and its noise of variance σ is $\eta(\sigma)$, then $\mathbf{x}(\mathrm{t})$ is encoded by the temporal spike pattern $\mathrm{a}(\mathrm{t})$ across the group of neurons governed by this equation:

$$\sum_n \delta(t - t_{in}) = G_i\left[\alpha_i \widetilde{\mathbf{\Phi}} \cdot \mathbf{x}(\mathrm{t}) + J_i^{bias} + \eta_i(\sigma)\right] \tag{1}$$

For decoding the spiking pattern as the vector $\hat{\mathbf{x}}(\mathrm{t})$, we need the linearly optimal decoding vector for $\mathbf{\Phi}$ each neuron.

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{\Gamma^{-1}\Upsilon} \\ \mathbf{\Gamma}_{\mathrm{ij}} &= \int \mathrm{a_i a_j} \mathbf{dx} \\ \mathbf{\Upsilon}_{\mathrm{j}} &= \int \mathrm{a_i} \mathbf{x} \mathbf{dx} \end{aligned} \tag{2}$$

The decoding vectors are the weighted with the post-synaptic current $h(t)$ induced by each spike:

$$\begin{aligned} \hat{\mathbf{x}}(t) &= \sum_i \mathrm{a}_i\big(\mathbf{x}(\mathrm{t})\big)\mathbf{\Phi}_{\mathrm{i}} \\ \mathrm{a}_i\big(\mathbf{x}(\mathrm{t})\big) &= \sum_n \delta(t - t_{in}) * h_i(t) \end{aligned} \tag{3}$$

We now have a neural group X which is representing $\mathbf{x}(\mathrm{t})$. We can also have another neural group Y which is representing $\mathbf{y}(\mathrm{t}) = \mathbf{f}\big(\mathbf{x}(\mathrm{t})\big)$ where $\mathbf{f}$, in the most general case, is a nonlinear transformation. We can generalize the derivation of the decoding vector to estimate the desired function $\mathbf{f}$. The transformational decoders for the neurons in Y such that their spiking pattern could be decoded to $\mathbf{y}(\mathrm{t})$ is derived by:

$$\boldsymbol{\Phi}^{\mathbf{f}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Upsilon}^{\mathbf{f}}$$

$$\boldsymbol{\Gamma}_{\mathrm{ij}} = \int \mathrm{a_i a_j}\mathbf{dx} \tag{4}$$

$$\boldsymbol{\Upsilon}_{\mathrm{j}}^{\mathbf{f}} = \int \mathrm{a_i}\mathbf{f}(\mathbf{x})\mathbf{dx}$$

Finally, let's consider a dynamical system characterized by:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(\mathrm{t}) + \mathbf{B}\,\mathbf{u}(\mathrm{t}) \tag{5}$$

We can represent this system within NEF by a recurrently connected neural ensemble:

$$\sum_n \delta(t - t_{in}) = G_i\big[\alpha_i \widetilde{\boldsymbol{\Phi}} \cdot \big(h_i(t) * \big[\acute{\mathbf{A}}\,\mathbf{x}(\mathrm{t}) + \acute{\mathbf{B}}\,\mathbf{u}(\mathrm{t})\big]\big) + J_i^{bias} + \eta_i(\sigma)\big]$$

$$\acute{\mathbf{A}} = \tau\,\mathbf{A} + \mathbf{I} \tag{6}$$

$$\acute{\mathbf{B}} = \tau\,\mathbf{B}$$

1.5 Moving from the literature to our computational models

A gaze-shift is planned in space, i.e. it is highly dependent on the percept of space. Space perception is constructed based on information received through multiple senses, and also on the internal processes of reasoning and associations, e.g. causal inference. Therefore, causal inference in spatial perception of audiovisual target(s) and planning of gaze-shifts towards such target(s) are interconnected processes. The question is how we can explain these various perceptual and motor phenomena in a common framework: 1) which is neurophysiologically plausible, 2) which satisfies the constraints on a proper cognitive architecture, 3) whose internal structure is able to account for sensorimotor, and possibly

cognitive, gaze-shift planning, 4) which is able to reproduce the experimental evidence on spatiotemporal variabilities of the causal inference and saccadic reaction time, 5) which is able to account for the known evidence for kinematics of 3D eye-head coordination.

Based on the presented arguments and evidence, in order to understand multisensory processing in neural and behavioral levels concretely, we need to have a large-scale model of functionalities and connectivities of cortical brain areas with the superior colliculus, basal ganglia, and the brainstem. In this thesis we provide computational-level descriptions for the representations and transformations within the proposed networks of expert information processing units. Represented signals and transformations for most expert units are inspired and supported by neurophysiological evidence.

In the second chapter, we suggest a model of how the brain uses the spatial and temporal features of the visual and auditory stimuli to infer whether or not they belong to a same object in the environment. The model includes controlled, leaky integrators to retain and process the transient sensory information. An evidence-accumulation decision-making circuitry is proposed to represent all possible scenarios, and selecting one based on constructing a spatiotemporal similarity measure. The result of the decision is implemented by selective disinhibition of plan representations through a striato-cortical projection. The model accounts for variability of reports of sameness when spatial and temporal distances between the stimuli change.

In the third chapter, we build upon the decision-making framework we previously used to solve the causal inference problem. Once a winning motor plan has been chosen based on causal inference, an instantaneous measure of confidence is computed based on the relative saliency of the winning motor plan compared to the alternate plans. A winning plan is only initiated when enough evidence is accumulated in its favor. This is realized by introducing an accumulative measure of confidence which integrates the instantaneous measure through time. A threshold is then set on the accumulative confidence and a GO command is released whenever it reaches the threshold. The model accounts for variability of saccadic reaction time as functions of spatial, temporal and reliability features of cross-modal stimuli.

In the fourth chapter, we propose a kinematic model of three-dimensional, head-unrestrained gaze-shifts. The input to this model is the eye-centered position of the target of the gaze-shift, supposedly provided by the superior colliculus, the output of the reaction time model. This target may be constructed based on either visual or auditory information, or integration of the two. The model achieves the gaze-shift through a spatiotemporal coordination of a saccadic eye movement, a head movement, and a vestibule-ocular eye movement. The model provides internal transformations that, together, account for behavioral evidence about gaze accuracy, relative eye and head contributions to gaze, non-commutivity of 3D rotations, Donders' law, Listing's law, and Fick constraints.

Each model is proposed within a "signals and systems" (Karris, 2003) framework. Every signal is a time-varying vector. Transformations of a signal is modeled in its feedforward conversions to other signals. Temporal processing of a signal can be controlled by its characterizing set of differential equations. Therefore, at a computational level, we have a network of signals and systems that characterize their behaviors. The whole network is designed to be neurally implemented in a circuitry of spiking neural networks, using the Neural Engineering Framework. Each signal is represented by a population of spiking neurons. Each representation can be transformed, according to a definite function, using its feedforward synaptic connections to other neural populations. Temporal processing of a representation is controlled by its recurrent connections between its neurons.

2 Causal Inference for Cross-Modal Action Selection: A Computational Study in a Decision Making Framework

Mehdi Daemi[1, 2, 3, 4], Laurence R Harris[1, 2, 4, 5], J. Douglas Crawford[1, 2, 3, 4, 5, 6 *]

1 Department of Biology and Neuroscience Graduate Diploma, York University, Toronto, ON, Canada

2 Centre for Vision Research, York University, Toronto, ON, Canada

3 Canadian Action and Perception Network

4 Department of Psychology, York University, Toronto, ON, Canada

5 School of Kinesiology and Health Sciences, York University, Toronto, ON, Canada

6 NSERC CREATE Brain in Action Program, York University, Toronto, ON, Canada

Front. Comput. Neurosci. **10**:62. doi:10.3389/fncom.2016.00062

Received: 21 Jan 2016; **Accepted:** 09 Jun 2016.

Edited by:

Concha Bielza, Technical University of Madrid, Spain

Reviewed by:

Sergey M. Plis, United States

Joaquin Goñi, Indiana University, USA

* **Correspondence:**
Dr. J. Douglas Crawford,
Center for Vision Research
Room 0009, Lassonde Bldg.
York University
4700 Keele Street
Toronto, Ontario, Canada, M3J 1P3
jdc@yorku.ca

2.1 Abstract

Animals try to make sense of sensory information from multiple modalities by categorizing them into perceptions of individual or multiple external objects or internal concepts. For example, the brain constructs sensory, spatial representations of the locations of visual and auditory stimuli in the visual and auditory cortices based on retinal and cochlear stimulations. Currently, it is not known how the brain compares the temporal and spatial features of these multisensory representations to decide whether they originate from the same or separate sources in space.

Here, we propose a computational model of how the brain might solve such a task. We reduce the visual and auditory information to time-varying, finite-dimensional signals. We introduce controlled, leaky integrators as working memory that retains the sensory information for the limited time-course of task implementation. We propose our model within an evidence-based, decision-making framework, where the alternative plan units are saliency maps of space. A spatiotemporal similarity measure, computed directly from the unimodal signals, is suggested as the criterion to infer common or separate causes.

We provide simulations that 1) validate our model against behavioral, experimental results in tasks where the participants were asked to report common or separate causes for cross-modal stimuli presented with arbitrary spatial and temporal disparities. 2) Predict the behavior in novel experiments where stimuli have different combinations of spatial, temporal, and reliability features. 3) Illustrate the dynamics of the proposed internal system. These results confirm our spatiotemporal similarity measure as a viable criterion for causal inference, and our decision-making framework as a viable mechanism for target selection, which may be used by the brain in cross-modal situations. Further, we suggest that a similar approach can be extended to other cognitive problems where working memory is a limiting factor, such as target selection among higher numbers of stimuli and selections among other modality combinations.

2.2 Introduction

Sensory systems detect different types of signals originating from objects in the surrounding environment. For example, visual information is carried by electromagnetic waves with a specific range of frequencies, whereas auditory information is carried by mechanical waves with a certain range of frequencies. Our brain constructs various perceptions and plans various actions in space and time, which can be triggered by sensations from multiple modalities. Integration of multimodal sensory information has been studied for temporal perceptions, e.g. perception of duration (Burr et al., 2009, Klink et al., 2011) and simultaneity (Harrar and Harris, 2008, Virsu et al., 2008), for spatial perceptions, e.g. spatial localization (Alais and Burr, 2004) and motion direction perception (Sadaghiani et al., 2009), for causal inference (Slutsky and Recanzone, 2001, Wallace et al., 2004), and also for action (Frens et al., 1995, Van Wanrooij et al., 2009). Here we are concerned with how the multisensory information is processed for causal inference.

Causal inference in animals is the process of estimating what events in outside world has caused the sensory representations in the brain (Shams and Beierholm, 2010, Lochmann and Deneve, 2011). In presence of multiple sensory representations, we compare their features to infer if they have a unique cause or not. A commonly studied case is when visual and auditory information is used to construct spatial and temporal perceptual features. If the temporal features are similar to each other, a common cause may be perceived overriding mismatches in their spatial features (Vroomen et al., 2001a, b, Godfroy et al., 2003). Similarly, if the spatial features are similar to each other, a common cause may again be perceived despite mismatches between temporal features (Vroomen and Keetels, 2006, Vroomen and Stekelenburg, 2011). These spatial and temporal binding effects break down at large spatial or temporal disparities (Slutsky and Recanzone, 2001, Wallace et al., 2004). In this paper we intend to propose a unique mechanism for causal inference which explains all this seemingly disparate evidence. However, let's first review some previous attempts on solving this problem.

In one study (Alais and Burr, 2004), observers were asked to report the location of a stimulus consisting of a flash and click presented with a spatial conflict. The spatial reliability of the visual signal was varied. The participants were told that the flash and click

belonged to a unique object. For the case of the most conspicuous visual stimuli they observed the classical ventriloquist effect such that the participants perceive the object close to the position of the visual stimulus. For heavily blurred visual stimuli, they perceive the object close to the auditory stimulus. For intermediate levels of blurriness, they perceive the object somewhere between the positions of the visual and auditory stimuli. Their results imply that, when the observers assume a common cause for the cross-modal stimuli, an intermediate position closer to the more reliable of the stimuli, is perceived as the location of the common cause. This idea was modeled, assuming Gaussian distributions for the unisensory cues, by Bayesian integration of the distributions, leading to the average of the two position cues weighted by the inverse of the variances of their distributions (Alais and Burr, 2004). Others tried to implement this optimal integration by a single-neuron model (Patton and Anastasio, 2003) or a model of a population of neurons (Ma et al., 2006).

Other experimental studies let the participants decide whether two cross-modal signals belonged to a unique object or not (Slutsky and Recanzone, 2001, Wallace et al., 2004). Such studies changed the spatial and temporal relationships between the two stimuli. For very short-duration and synchronous stimuli, the participants reported a unique object as the source of the signals and perceived it at the weighted average of the position of the two signals. When the presentation time was extended or temporal disparity was introduced between the signals, the chance of reporting a unique cause for two spatially disparate signals decreased drastically. Also for synchronous stimuli, increasing the spatial disparity between the stimuli decreased the percentage of the trials in which the participants reported a unique object as the source. Their results showed that when participants are not told to assume a common source for the stimuli, they might localize the stimuli in common or separate spatial positions depending on the spatial and temporal features of the stimuli.

Some theoretical studies have tried to model the effect of spatial disparity (Hairston et al., 2003) on the report of a common cause (Kording et al., 2007, Sato et al., 2007). However, these studies ignored the temporal effect. They used the uncertain spatial cues, detected through multiple sensory channels, to calculate the probabilities of them arising from same or separate sources. If the same source is more likely, these models calculate the optimal estimate of the location of the same source as the weighted average of the cues. If separate

sources are more likely, the models shown that the uncertain spatial cues are the best estimates of the two locations. A physiologically realistic framework for these models has not been offered (Ma and Rahmati, 2013). Some other theoretical studies reduce the criterion for fusion to the temporal features of the events, ignore the spatial disparity, and propose that the cross-modal events are bound together if they happen within a relative time window (Colonius and Diederich, 2010, Diederich and Colonius, 2015).

Here we want to propose a more general approach which considers the spatial and temporal dimensions in a common framework. We suggest a model of how the brain solves the causal inference problem for spatial localization for cross-modal, audiovisual stimuli with arbitrary spatial and temporal disparities. We propose the model at the computational level (Marr, 1982), not assuming a specific probability distribution or neural representation for the spatial position of the stimuli. We consider two stimuli, visual and auditory, with only spatial and temporal features. However, other problems with more than two stimuli, or with other modality combinations, or with stimuli of semantic or emotional significance can also be tackled by our approach. We consider the stimuli to be composed of multidimensional, time-varying, position signals which communicate the time and place of the stimuli. Our model is proposed within an evidence-based decision making framework including a short-term memory, in the form of a leaky integrator, and a spatiotemporal similarity measure as the criterion for inferring the cause of the input signals. The short-term memory retains spatial information (not information about the order and temporal interrelations of events) and our similarity measure captures spatial and temporal disparities between the stimuli (not a higher-level order relation between them in time or space). We use this model to simulate known psychophysical results, and to generate predictions that can be used to test the model. Such results demonstrate that a model constructed in a decision making framework and inferring a causal structure based on a spatiotemporal similarity measure explains the behavioral results and could possibly be used by the brain to solve the target selection problem when cross-modal stimuli are presented.

2.3 Model Overview

The problem we are addressing is causal inference in localization of cross-modal stimuli in which the spatiotemporal properties of the components vary. To solve this problem, we borrow two popular concepts from cognitive neuroscience that (perhaps surprisingly) have not yet been incorporated into models of multisensory spatial integration: decision making (Wang, 2008), and working memory (Baddeley, 2003b). Although the computations in this model could pertain to any cognitive or behavioral use of causal inference from multimodal inputs, we designed this model with output to the gaze control system in mind, because this is one of the best understood systems in the brain (Bell et al., 2005) and because numerous gaze-control laboratories are capable of testing our predictions. Thus, one can think of the output of the model as dictating whether a gaze-shift will be made toward the visual stimulus, the auditory stimulus, or a combined representation derived from both. Finally, we have arranged the general order and nature of our model algorithms to be compatible with the known biology of these systems but focus the current study on replicating and predicting psychophysical results.

The sensory information received from stimuli in the environment is transient as most stimuli are only present for a limited time. Sensory information about the position and reliability of multimodal stimuli is moved to, and temporarily stored in, working memory where operations such as integration and computation of similarity take place. Working memory, in general, is used to bring together different pieces of information for cognitive processing with the goal of performing tasks such as reasoning, problem solving or action planning (D'Esposito et al., 1995, Baddeley, 2003a). Working memory is a distributed system in the brain, with multiple brain areas activated depending on the specific task being implemented (Courtney et al., 1997, Haxby et al., 2000, Fuster, 2004). Working memory in our model comprises four computational units (shown in blue in figure 1) that are responsible for retaining sensory information, integrating spatial cues, computing a similarity measure, and feeding the decision-making circuitry (Bechara et al., 1998).

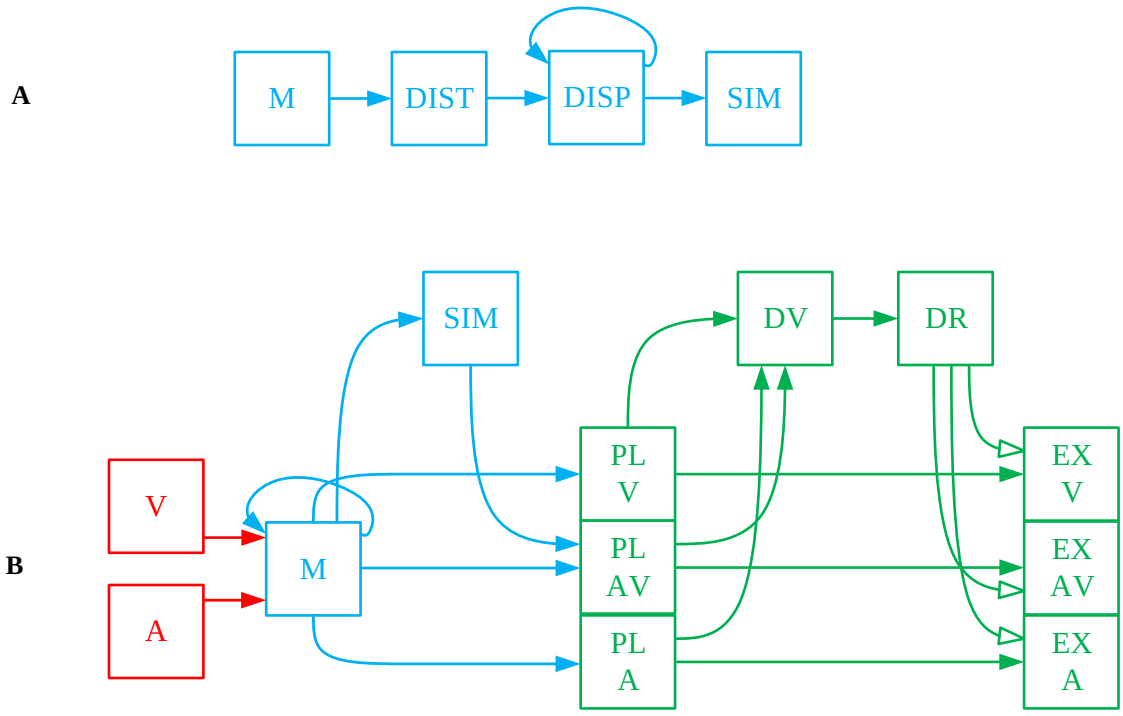

Figure 1: **A) Computation of the spatiotemporal similarity measure and using it to make a call on the sameness of the cause of cross-modal signals.** The eye-centered, spatial components of the visual and auditory signals, which are stored in short-term memory (M) are fed into the unit DIST to calculate the spatial distance between them as a function of time. The spatial distance is then sent to the unit DISP, called spatiotemporal disparity, where it is integrated across time. The spatiotemporal disparity is then sent to the unit SIM, called spatiotemporal similarity measure, where is goes through an inverting and normalizing function. The spatiotemporal similarity measure is used for making a call on the sameness of the cause of the two received signals. **B) The complete model of gaze-shift, target selection in cross-modal situations.** The visual (V) and auditory (A) signals are stored in a multisensory memory (M) structure. In parallel, the visual and auditory signals are used for computation of spatiotemporal similarity measure (SIM), as illustrated in detail in **A**. The three alternative plans are constructed as saliency maps from the memorized information and are represented initially in the plan layer in three units PL_V, PL_AV, PL_A. The unisensory plans are the unisensory stimulus positions along with their reliabilities which are regarded as equivalent to their saliencies. The multisensory plan is the weighted average (by reliabilities) position of the cross-modal stimuli along with the similarity measure as its saliency. The decision variable is constructed in the unit DV by communicating the saliencies of the three plans. The result of the decision is computed in DR by a function which implements the idea that the multisensory plan wins if the similarity measure is greater than a threshold and the more reliable of the unisensory plans wins if the similarity measure is lower than the threshold. The spatial components of the three plans are communicated from the plan layer to three units EX_V, EX_AV, EX_A in the execution layer. The result of the decision is materialized by selective inhibition of the plan representations in the execution layer. Only the winning plan is disinhibited, based on the decision result, and is sent for execution.

We propose our model within a decision-making framework. Decision making is the process of deliberation resulting in the commitment to one of multiple alternative plans (Gold and Shadlen, 2007, Heekeren et al., 2008, Cisek and Kalaska, 2010). The deliberative process consists in the accumulation of evidence through processing the available information. This is realized in the evolution of systemic decision variables through time. The result of the decision is determined by a rule which is applied to the decision variables. Decision rules determine how or when the decision variable is interpreted to arrive at a commitment to a particular plan (Churchland et al., 2008). The decision result is the output variable of the evidence accumulation and rule application, that determines which plan is to be executed. Accumulation of evidence changes the decision variables and may change the decision result (Bogacz, 2007). As we shall see, each of these features has been incorporated into our model (green in figure 1B).

The first part of the decision is to decide whether there is a unique cause for the two signals or if they correspond to two separate events. As explained before, the experimental evidence shows that this decision is determined based on the spatial and temporal relationship between the cross-modal stimuli (Wallace et al., 2004). We propose a measure of spatiotemporal similarity between the two received signals that is used for making this decision. Figure 1A shows how this measure is calculated in working memory. The spatiotemporal pattern of stimuli presentation is captured in a temporally changing spatial position signal, decoded from the representations of sensory space in the brain. Spatial distance ($DIST$) between the two stimuli, as a function of time, is first calculated. Spatial distance is integrated through time to calculate the spatiotemporal disparity ($\overrightarrow{DISP}$). Spatiotemporal similarity measure (SIM) is calculated by applying a function that inverts and normalizes the spatiotemporal disparity. This time-varying, similarity measure decreases with increases in the spatial disparity and / or temporal disparity between the two presented stimuli.

The complete problem can be conceptualized as choosing between three possible scenarios: 1) the signals are coming from one same object. In this case the target for gaze-shift is constructed as a weighted average of the visual and auditory estimates. 2) The signals are coming from different objects and the visual stimulus is more salient, in which case the

target is chosen to be at the location of the visual stimulus. 3) The signals are coming from different objects and the auditory stimulus is more salient, so, the target is chosen to be at the location of the auditory stimulus. Thus, the main task for our model is to infer one of these three scenarios from a given pair of multisensory inputs.

The complete model is shown in figure 1B. The inputs to the system are the temporally changing position signals of the visual and auditory stimuli along with their reliabilities ($\vec{V}$ and $\vec{A}$). These spatial position signals are temporarily stored in a memory structure ($\vec{M}$). The spatiotemporal similarity measure (SIM) is computed from the position signals stored in memory. The previously mentioned three possible scenarios are physically realized in the form of three plan representations. Each plan unit represents the potential goal for an attention shift (if that plan wins) along with the saliency of the plan. The visual ($\overrightarrow{PL\_V}$) and auditory ($\overrightarrow{PL\_A}$) plan units represent the position of the corresponding stimuli along with their reliabilities (Kording et al., 2007, Rowland et al., 2007) (as our stimuli don't bear any emotional significance or semantic meaning, their saliency is reduced to their reliability). Reliability in our model is a one-dimensional, real-valued parameter, which can change between 0 and 1 for the least to most reliable, and is an input to the model. We presume that this reliability can be calculated, upstream of our model, based on the representation of the spatial position, e.g. the inverse of the variance for a normal distribution (Kording et al., 2007, Ohshiro et al., 2011). The multisensory plan ($\overrightarrow{PL\_AV}$) unit represents average of the positions of the two stimuli weighted in proportion to their respective reliabilities (Alais and Burr, 2004). The saliency of the multisensory plan is proposed to be the spatiotemporal similarity measure.

The decision variable ($\overrightarrow{DV}$) is constructed from the saliencies of the three alternative plans. The decision on same or separate causes for the signals is made by comparing the saliency of the multisensory plan with a threshold. We assume this threshold is tunable, and one possible way to account for the effects of emotional or semantic value of stimuli on sensory fusion is to be able to adjust this threshold. However, as this is beyond the scope of our model, we set the threshold to 0.5 (to match the experimental evidence, see below) and for consistency we use the same value for all of our predictive simulations. As long as saliency,

i.e. the spatiotemporal similarity measure, is above threshold the decision that they are from the same source is preferred. If the similarity measure drops below threshold the decision changes to that they originate from separate sources. In this case, the decision concerning which cause forms the goal of a shift of attention is made by comparing the saliencies of the two unisensory plans. The overall result of this three-way decision ($\overrightarrow{DR}$) is stored as a 3-D signal that allows communication of only the winning plan to the execution units ($\overrightarrow{EX\_V}$, $\overrightarrow{EX\_A}$, $\overrightarrow{EX\_AV}$). This is implemented through the decision result. $\overrightarrow{DR}$ keeps all EX units under constant inhibition. When a plan wins, its corresponding EX unit is disinhibited.

The general outline of the model is inspired by known properties of the visual, auditory, and gaze control systems. The visual signal is the position of the visual stimulus in eye-centered coordinates (Andersen et al., 1997, Maier and Groh, 2009). Auditory space is encoded initially in a craniocentric frame of reference (Knudsen and Konishi, 1978, Knudsen and Knudsen, 1983) as the auditory receptors are fixed to the head. For multisensory information processing and motor planning, the two sensory signals, $\vec{V}$ and $\vec{A}$, should be in a common reference frame (Jay and Sparks, 1987, Andersen et al., 1997) which has been shown to be eye-centered for action involving the gaze-control system and early aspects of reach planning (Groh and Sparks, 1992, Cohen and Andersen, 2000, Pouget et al., 2002). The sensory signals are then sent to the distributed network of working memory. Posterior parietal and dorsolateral prefrontal cortex have been shown to actively maintain such signals (Funahashi et al., 1989, Cohen et al., 1997), similar to the short-term memory $\vec{M}$ in our model. The prefrontal cortex is involved in the higher-order, executive functions of working memory (D'Esposito and Postle, 2015), including integration of the signals into unique events, realized in our model through $\overrightarrow{DIST}$, $\overrightarrow{DISP}$, and $\overrightarrow{SIM}$. It is thought that the working memory then feeds the plan representations of the decision making circuitry in frontal cortex (Jones et al., 1977, Canteras et al., 1990, Berendse et al., 1992, Yeterian and Pandya, 1994, Levesque et al., 1996), like our plan representations in plan layer PL. Plan representations are then thought to send bids, e.g. their saliencies as in our case, to a central arbitrating system (Redgrave et al., 1999), e.g. the telencephalic decision centers, that gate their access to effectors. This is represented in our model through $\overrightarrow{DV}$ and $\overrightarrow{DR}$ and their connection which realizes a decision rule. The basal ganglia are

thought to receive the result of the decision from cortex (Beiser and Houk, 1998, Koos and Tepper, 1999, Gernert et al., 2000) and implement it through selective disinhibition of cortical channels, which is abstracted in our model through the multiplicative effect of the $\overrightarrow{DR}$ on plan representations in execution layer EX. In order to plan a gaze-shift, for example, the final winning plan is sent to the superior colliculus (Munoz and Guitton, 1989, Klier et al., 2001). This command could then be used to drive the eye-head coordination system (Klier et al., 2003, Daemi and Crawford, 2015) to reorient the line of sight to the appropriate target.

2.4 Mathematical Formulation

2.4.1 Method

Our model implements causal inference through a decision making network for planning actions in a dynamic environment. This contrasts to previous approaches which either described 1) inference as chains of if-then rules which statically transform the internal states of the system (Newell and Simon, 1972, Anderson, 1983) or 2) goal-directed motor planning within the time constraints of environmental interactions (van Gelder, 1998). While the former approach ignores the short-term dynamics of perception and action, the latter ignores the internal system, and sacrifices the high-level linguistic processes, such as complex planning and deductive reasoning. Our goal was to integrate both "dynamic perception / action" and "high-level inference" in a way consistent with our knowledge of human and animal cognitive systems (see section 3).

To do this, we adopt a unified approach where a model is identified by functions of both the internal state variables and the time. Inspired by the brain, such more general models are realized through a distributed network of parallel processing units. This approach simultaneously accounts for syntactic manipulations of representations underlying inference, and flexible control of information routing between different units through time (Eliasmith, 2013). Although we do not deal here with the neural implementation of the model, all the representations and transformations are designed based on the known neurophysiology, and can be neurally realized by a recent theoretical approach, neural engineering framework, which unifies the symbolic, connectionist, and dynamicist viewpoints (Eliasmith and Anderson, 2003, Eliasmith et al., 2012). The relatively high

number of variables in such models is because we are modelling an adaptive, robust biological system which can behave and survive in an uncertain, changing environment.

More specifically, we implement an evidence-based decision making process, whose representations are evolving through time. The inference's syntactic manipulations are realized through selective inhibition of plan representations, as inspired by the brain. Routing the information through the system is realized in a unified architecture where all attractor networks are controlled integrators which include a dimension (controlled leak) whose value controls whether the structure updates its value by it input, retains its current value, or clears its content. Information routing is controlled by the dynamics of the system not by the choice of modeler, as it is in the brain. As a result, inference is realized through time, evolving as empirical evidence is accumulated, helping us to survive in a highly dynamic environment.

2.4.2 Unisensory Signals

When visual or auditory stimuli occur in the environment, they are percieved at specific spatial locations, within specific time windows. The visual stimulus is encoded in retinal coordinates, i.e. an eye-centered frame of reference. The auditory stimulus is initially encoded relative to head, however, for cognitive and motor purposes, this code is transformed into an eye-centered reference frame as well (Maier and Groh, 2009). The unisensory input signals in our model are transient, time-varying, four-dimensional vectors. The four dimensions include a first component for existence of the signal, a second component for reliability of the signal and two last components for the eye-centered position of the signal in the spherical coordinates. The existence component gets value 1 or 0 based on whether or not a stimulus is detected in the environment, by stimulation of the sensory receptors. It controls the interaction of the sensory information with memory (explained next). The reliability component, changing between 0 and 1 for least to most reliable, is computed from the early representation of the signal (Kording et al., 2007, Ohshiro et al., 2011), and indicates how reliable the representation is about the position of the stimulus.

$$\vec{V}(t) = \begin{bmatrix} ext_v \\ rel_v \\ ech_v \\ ecv_v \end{bmatrix} \quad (1)$$

$$\vec{A}(t) = \begin{bmatrix} ext_a \\ rel_a \\ ech_a \\ ecv_a \end{bmatrix} \quad (2)$$

2.4.3 Short-Term Memory

The transiently presented sensory signals need to be temporarily stored for further cognitive processing, e.g. inference (D'Esposito et al., 1995, Baddeley, 2003a), and then feeding the decision making circuitry. Accordingly, the unisensory signals are first communicated a short-term memory structure. It is a state space of finite dimensions which temporarily stores the unisensory signals in a unique representation. It consists of leaky integrators with controllable leaks. Sensory information is retained across eight dimensions of this state space, four dimensions for each modality. Those four modality-specific dimensions include a first component controlling the integrator's leak, and three components storing the last three dimensions of the unisensory signals:

$$\vec{M}(t) = \begin{bmatrix} lk_{mv} \\ rel_{mv} \\ ech_{mv} \\ ecv_{mv} \\ lk_{ma} \\ rel_{ma} \\ ech_{ma} \\ ecv_{ma} \end{bmatrix} \quad (3)$$

This memory structure, in connection with the transient sensory signals, is governed by these nonlinear state-space equations. In a general sense, such state space equations are the basis of constructing attractor neural networks which is believed to underlie memory structures in the brain (Conklin and Eliasmith, 2005, Singh and Eliasmith, 2006). The boundary and input conditions of these differential equations are dictated by a dynamic environment. Therefore, the current state of the state space is controlled internally, by the

controllable leaks, in constant interaction with the environment. However, more specifically, before any input comes in, all dimensions of the state space are zero.

$$\begin{bmatrix} \dot{lk}_{mv} \\ \dot{rel}_{mv} \\ \dot{ech}_{mv} \\ \dot{ecv}_{mv} \\ \dot{lk}_{ma} \\ \dot{rel}_{ma} \\ \dot{ech}_{ma} \\ \dot{ecv}_{ma} \end{bmatrix} = \begin{bmatrix} 0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,(1-lk_{mv})0\,0\,0\,0\,0\,0 \\ 0\,0\,(1-lk_{mv})0\,0\,0\,0\,0 \\ 0\,0\,0\,(1-lk_{mv})\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,(1-lk_{ma})0\,0 \\ 0\,0\,0\,0\,0\,0\,(1-lk_{ma})0 \\ 0\,0\,0\,0\,0\,0\,0\,(1-lk_{ma}) \end{bmatrix} \times \begin{bmatrix} lk_{mv} \\ rel_{mv} \\ ech_{mv} \\ ecv_{mv} \\ lk_{ma} \\ rel_{ma} \\ ech_{ma} \\ ecv_{ma} \end{bmatrix} + \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix} \times \begin{bmatrix} ext_v \\ rel_v \\ ech_v \\ ecv_v \\ ext_a \\ rel_a \\ ech_a \\ ecv_a \end{bmatrix} \tag{4}$$

The controllable leaks characterize the behavior of the controlled integrator (table 1) (Eliasmith, 2005). The two leaks are fed by the existence component of the corresponding sensory input. The existence component is 1 when the stimulus is present and is 0 when it is not, so the leaks always assume digital values 0 or 1. This means the integrator is updated by the new input when the input is present and maintains the current value when no input is present.

	Leak = 0	Leak = 1
No input coming	Keeps the current value	Clears the memory
Input coming	Integrates and accumulates the input	Updates to the input

Table 1: the effect of the leak on the behavior of a leaky integrator. Theoretically speaking, if the leak gets a value between 0 and 1, when there is no input the integrator clears the memory with a speed controlled by the leak, and when there is an input the integrator integrates the input with a speed controlled by the leak. However, both our integrator structures always assume digital values of 0 or 1.

2.4.4 Spatiotemporal Similarity Measure

The cognitive processing in working memory, in our model, consists of computing a measure of similarity between the two unisensory signals based on their spatial positions and temporal profiles. Figure 1A illustrates the connectivity of structures for calculating this measure. We start with the spatial distance $DIST$. The spatial distance between the two unisensory stimuli is calculated from the information stored in the short-term memory about the spatial positions of the stimuli. It is computed, in spherical coordinates, in the connection from $\vec{M}$ to $DIST$:

$$\begin{aligned} DIST(t) = \ & [dist] \\ = \ & \cos^{-1}[\cos(\ ech_{mv}) \times \cos(\ ech_{ma}) \\ + \ & \sin(\ ech_{mv}) \times \sin(\ ech_{ma}) \times \cos(\ ecv_{mv} - \ ecv_{ma})] \end{aligned} \quad (5)$$

The spatiotemporal disparity $\overrightarrow{DISP}$ is then calculated from the spatial distance by integrating it across time. Our proposed structure is a state space of two dimensions. This is, again, a leaky integrator with controllable leak. The two dimensions of this state space include a first component controlling the integrator's leak and a second component where the integrated value of the spatial distance is accumulated.

$$\overrightarrow{DISP}(t) = \begin{bmatrix} lk_{disp} \\ disp \end{bmatrix} \quad (6)$$

These state space equations characterize the behavior of this integrator. Before introduction of inputs, all dimensions of the state space are zero.

$$\begin{bmatrix} \dot{lk_{disp}} \\ \dot{disp} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & (1 - lk_{disp}) \end{bmatrix} \times \begin{bmatrix} lk_{disp} \\ disp \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \times [dist] \quad (7)$$

Here, the leak does not need to be controlled based on existence of the input. The leak is internal to the functioning of the integrator, and represents a value 0 all through the stimulus presentation window. That is because we want it to integrate the input when there is any, and retain the current value when there is no input (table 1). The result of this integration gives us a measure of spatiotemporal disparity between the visual and auditory stimuli. A tangent hyperbolic function is then applied on the disparity measure to calculate a measure of similarity between the two stimuli:

$$SIM(t) = [sim] = 1 - tanh(0.5 \times disp) \quad (8)$$

This makes the similarity measure change between 0 and 1 for the least to the most similar. Equations in this section might not be supported by a known brain mechanism, however, we will later show that using spatiotemporal similarity as the criterion to infer unique or separate causes can explain the experimental evidence about the relation of such judgements with the spatial and temporal disparities between cross-modal stimuli.

2.4.5 Decision Making Process

The information processed in working memory is then communicated to the decision making circuitry (Bechara et al., 1998), which realizes the causal inference in our model. We introduce three plan units, visual, auditory and multisensory, which are fed by the working memory. Each of these channels is a 3-dimensional vector whose first component represents the saliency of that plan. The saliency of each of the unisensory plans is reduced to its reliability. The last two components of the two unisensory plans represent their respective spatial positions as stored in short-term memory:

$$\overrightarrow{PL\_V}(t) = \begin{bmatrix} sal_{plv} \\ ech_{plv} \\ ecv_{plv} \end{bmatrix} = \begin{bmatrix} rel_{mv} \\ ech_{mv} \\ ecv_{mv} \end{bmatrix} \quad (9)$$

$$\overrightarrow{PL\_A}(t) = \begin{bmatrix} sal_{pla} \\ ech_{pla} \\ ecv_{pla} \end{bmatrix} = \begin{bmatrix} rel_{ma} \\ ech_{ma} \\ ecv_{ma} \end{bmatrix} \quad (10)$$

Integration of the unimodal signals, which might be used to drive a gaze-shift, is implemented in working memory, in its connection to multisensory plan representation. The multisensory channel represents a weighted average of the positions of the two stimuli, weighted by their reliabilities. The saliency of the multisensory plan is considered to be the spatiotemporal similarity between the two stimuli, which varies between 0, for least similar, and 1, for most similar:

$$\overrightarrow{PL\_AV}(t) = \begin{bmatrix} sal_{plav} \\ ech_{plav} \\ ecv_{plav} \end{bmatrix} = \begin{bmatrix} sim \\ rel_{mv} \times ech_{mv} + rel_{ma} \times ech_{ma} \\ rel_{mv} \times ecv_{mv} + rel_{ma} \times ecv_{ma} \end{bmatrix} \quad (11)$$

Now, we are ready to construct our decision variable, realizing a central decision center (Gold and Shadlen, 2007). We propose a three-dimensional vector as the decision variable $\overrightarrow{DV}$ which is completely characterized by the saliency of the plan (PL) representations:

$$\overrightarrow{DV}(t) = \begin{bmatrix} dv_v \\ dv_a \\ dv_{av} \end{bmatrix} = \begin{bmatrix} sal_{plv} \\ sal_{pla} \\ sal_{plav} \end{bmatrix} \quad (12)$$

The values of the components of $\overrightarrow{DV}$ determine the decision about which of the visual, auditory or multisensory channels drives the final goal of gaze-shift. The result of this decision is to disinhibit the desired channel and keep inhibiting the undesired ones (explained below). The result of the decision making process is temporarily stored in another structure that we call 'decision result' or $\overrightarrow{DR}$. The decision function, which transforms $\overrightarrow{DV}$ to $\overrightarrow{DR}$, is the abstract underlying mechanism of inference in our model, and is formed through this idea:

$$Decision\ Result = \begin{cases} \begin{bmatrix}1\\0\\0\end{bmatrix} & if\ sim > threshold \\ \begin{bmatrix}0\\1\\0\end{bmatrix} & if\ sim < threshold\ and\ rel_v > rel_a \\ \begin{bmatrix}0\\0\\1\end{bmatrix} & if\ sim < threshold\ and\ rel_a > rel_v \end{cases} \quad (13)$$

Which is mathematically realized by this proposed functionality:

$$\overrightarrow{DR}(t) = \begin{bmatrix} dr_v \\ dr_a \\ dr_{av} \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{-sl_{av}(th_{av}- dv_{av})}} \times \frac{1}{1+e^{-sl_u(dv_v-dv_a)}} \\ \frac{1}{1+e^{-sl_{av}(th_{av}- dv_{av})}} \times \frac{1}{1+e^{-sl_u(dv_a-dv_v)}} \\ \frac{1}{1+e^{-sl_{av}(dv_{av}-th_{av})}} \end{bmatrix} \quad (14)$$

th_{av} is the tunable threshold for the similarity measure above which we perceive the two signals as coming from the same object and below which we can differentiate the cause of the two signals. sl_{av} and sl_u are function parameters which determine the speed and confidence of the transition between alternative decisions.

The decision result controls the communication of the plan representations from the plan layer, PL, to the execution layer, EX. Accordingly, the plan representations in EX are governed by:

$$\overrightarrow{EX\_V}(t) = \begin{bmatrix} ech_{exv} \\ ecv_{exv} \end{bmatrix} = dr_v \times \begin{bmatrix} ech_{plv} \\ ecv_{plv} \end{bmatrix} \quad (15)$$

$$\overrightarrow{EX\_A}(t) = \begin{bmatrix} ech_{exa} \\ ecv_{exa} \end{bmatrix} = dr_a \times \begin{bmatrix} ech_{pla} \\ ecv_{pla} \end{bmatrix} \quad (16)$$

$$\overrightarrow{EX\_AV}(t) = \begin{bmatrix} ech_{exav} \\ ecv_{exav} \end{bmatrix} = dr_{av} \times \begin{bmatrix} ech_{plav} \\ ecv_{plav} \end{bmatrix} \quad (17)$$

$\overrightarrow{DR}$ implements the decision concerning which plan drives the gaze-shift. This is applied by selective inhibition of plan representations in the execution layer (EX). EX plan representations are selectively inhibited to determine the winning plan. Here, this is shown

by the multiplicative effect of the corresponding $\overrightarrow{DR}$ component. Such functionality can be neurophysiologically realized by an inhibitory connection from a neural population representing $\overrightarrow{DR}$ to the neural populations representing the execution layer (EX) plans (Redgrave et al., 1999, Sajad et al., 2015).

2.5 Results

Psychophysicists record the observable behavior of subjects during experiments. However, the neurocognitive internal system underlying the behavior is not accessible to the psychophysicist. For example, for causal inference studies in cross-modal spatial localization, the "report of sameness" is the only measureable behavior, while the whole host of internal mechanisms, e.g. sensory representations, working memory and decision making units, which are responsible for the behavior are not measurable. In this paper we propose a model of the internal cognitive system underlying the implementation of such tasks. In this section: 1) we verify our model against the limited number of psychophysical studies of causal inference during cross-modal spatial localizations which systematically varied both the spatial and temporal features (Slutsky and Recanzone, 2001, Wallace et al., 2004). We do so (in 5-1) by comparing our model's output with the only measureable behavior "report of sameness" in such experiments. 2) At this stage, we have verified the ability of the model to reproduce the human behavior when the spatial and temporal configurations of the cross-modal stimuli are varied. We then look into the internal system by illustrating the dynamics of the decision variable and decision result when we change the spatial (5-2) or temporal (5-3) disparities between the stimuli. 3) We then use the model to predictively simulate the human behavior in some novel situations where experimental evidence is not yet available. We first simulate what happens when the reliability of the stimuli vary, when separate sources are perceived (5-4). Then we will illustrate how accumulation of evidence through exposure of the model to temporally extended stimulus presentations may change the decision (5.5).

2.5.1 Inference of a Unique Cause for Cross-Modal Stimuli

The percentage of the times that an audio-visual stimulus is judged as arising from a unique cause varies with the spatial and temporal features of the stimuli (Slutsky and Recanzone, 2001, Wallace et al., 2004). Slutsky and Recanzone (2001) kept the position, duration, and

onset of the auditory stimulus fixed, and varied the onset and position of the visual stimulus and found how this report of unique cause changes. They found that a unique cause was elicited for small temporal disparities even at large spatial disparities, and also for large temporal disparities for small spatial disparities (Slutsky and Recanzone, 2001).

Figure 2 shows the output of our model when stimulus parameters are varied in the same way as Slutsky and Recanzone (2001). Our proposed criterion for this decision is the measure of spatiotemporal similarity. This measure is shown as a function of temporal disparity for different spatial disparities in figure 2A and as a function of spatial disparities for different temporal disparities in 2B. The decision is made by applying a threshold (set to 0.5 throughout all of our simulations) function to the similarity measure: if it is above threshold, the decision is that there is a unique cause, if it is below threshold the decision is that there are separate causes. The results of this decision are shown in figure 2C as a function of temporal disparity for different spatial disparities and in 2D as a function of spatial disparities for different temporal disparities.

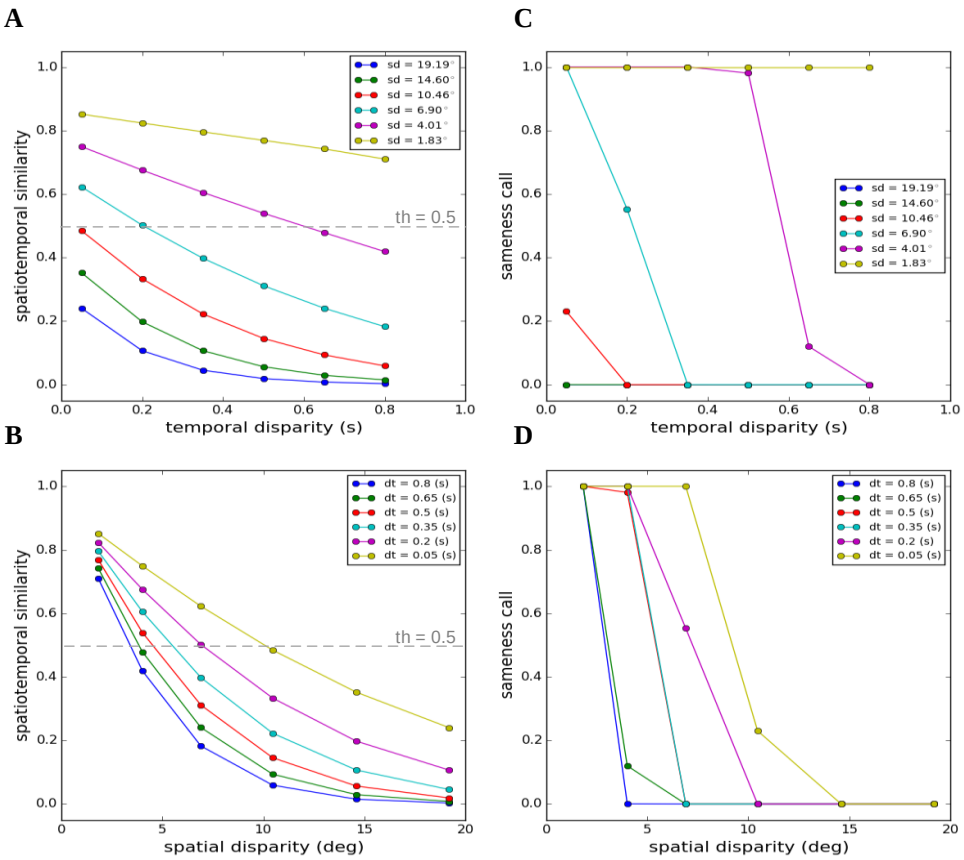

Figure 2: **Spatiotemporal similarity measure as the criterion for the decision on the uniqueness of the cause.** Here we replicate a task where participants were asked to report if two cross-modal stimuli emanated from a unique cause (Slutsky and Recanzone, 2001). While all features of the auditory stimulus were kept fixed, they systematically varied the spatial position and the onset time of the visual stimulus and studied how the sameness report changes. **A)** Spatiotemporal similarity measure as a function of temporal disparity for different spatial disparities. **B)** Spatiotemporal similarity measure as a function of spatial disparity for different temporal disparities. **C)** Sameness call as a function of temporal disparity for different spatial disparities. **D)** Sameness call as a function of spatial disparity for different temporal disparities. The values '1' and '0' for the sameness call indicate the same source and separate sources respectively. The symbols 'sd' and 'dt' indicate the spatial (degrees) and temporal disparities (seconds) respectively. The grey dashed lines in A and B indicate the threshold applied to the similarity measure.

The average percentage of the reports of a unique cause, among a number of participants and through multiple trials, changing by the spatial and temporal disparities, follow a meaningful pattern, as experimentally observed (Slutsky and Recanzone, 2001). This pattern is closely captured by the trends produced by our model which infers the causal structure based on the spatiotemporal similarity. Unique cause is predicted for a wide range of temporal disparities if the spatial disparity is very small, as shown in figure 2A and 2C for a spatial disparity of 1.83° (ventriloquism effect). The "sameness call" changes at some point for most spatial disparities if the temporal disparity becomes greater than threshold. Similarly, the "sameness call" changes for a given temporal disparity if the spatial disparity exceeds some threshold. Thus, although we did not tinker extensively with our model parameters to exactly match the experimental results quantitatively, we conclude that the model replicates the key results and principles of the published experiment.

2.5.2 Effect of Spatial Disparity

Spatial proximity is one of the features used to judge whether or not two signals have a common source (Hairston et al., 2003, Wallace et al., 2004). Figure 3 shows the performance of our model for a task in which visual and auditory stimuli have the same onset time (0.2 seconds) and duration (0.3 seconds). While the position of the visual stimulus was fixed, the position of the auditory stimuli was varied systematically (spatial disparities from 1.5° to 21.7°, figure 3A). The end behavior, "sameness call", of our model for this task has already been validated by experimental results in section 5-1, the yellow lines (very low temporal disparity) in figures 2-B and 2-D, and we want to show the internal dynamics here. Figure 3B shows the similarity measure, represented in the multisensory dimension of the decision variable, for the five spatial disparities. Figure 3C shows the "sameness call", represented in the multisensory dimension of the decision result, for each spatial disparity. For a fixed temporal structure, the similarity measure decreases when the spatial distance increases. There is a point, around 10° of spatial distance for this case, where the decision about the uniqueness of the cause changes. Our model proposes that the reason is that the similarity measure drops below threshold, and when this happens the unisensory plan with the higher saliency wins and is executed (not shown here). These simulations show how the temporal evolution of the internal system is influenced when the

spatial disparity between cross-modal stimuli varies, sometimes leading to a change in decision through time (sd = 15.6° or 21.7° here).

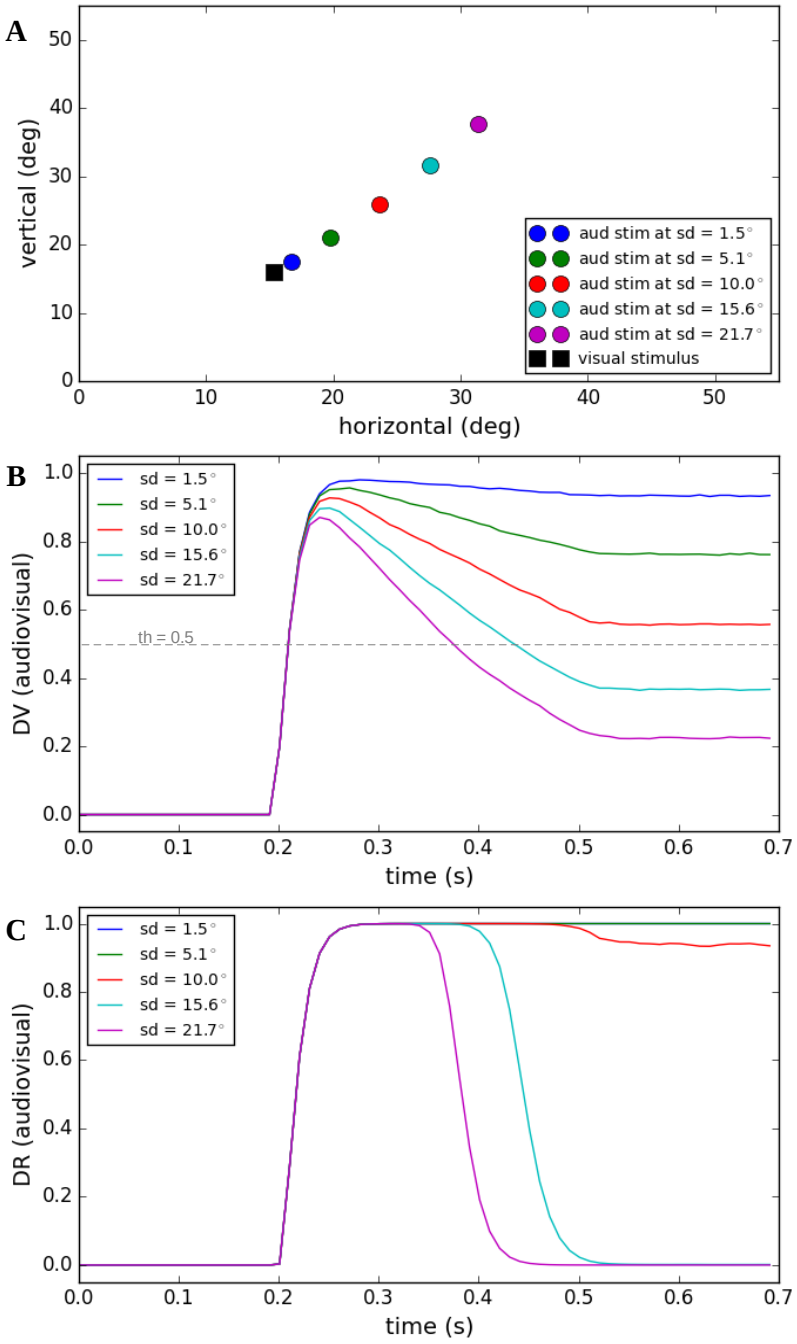

Figure 3: **Effect of spatial disparity of cross-modal stimuli on target selection.** Five different conditions have been considered (illustrated by color coding). The spatial and temporal features of the visual stimulus and the temporal features of the auditory stimulus are fixed for all conditions. The spatial position of the auditory stimulus changes in each condition. **A)** The spatial position of the visual stimulus and the five different spatial positions of the auditory stimulus, in the five conditions, are shown. **B)** The multisensory component of the decision variable is shown for different conditions as a function of time. It changes based on the spatial disparity of the stimuli in each condition. The unisensory components do not change. **C)** The multisensory component of the decision result is shown for the different conditions. It is one for smaller spatial disparities (indicating a common cause) and changes to zero (indicating separate causes) when the spatial disparity exceeds the threshold (shown as a dashed line in B).

2.5.3 Effect of Temporal Disparity

Temporal disparity is another feature that contributes to the decision about the sameness of the cause of the signals (Wallace et al., 2004, Chen and Vroomen, 2013). In figure 4 we show the simulations of our model under a task in which the visual and auditory stimuli have fixed positions close to each other. The duration of the auditory stimulus and visual stimulus are fixed (0.3 seconds). As shown in figure 4A, while the onset time of the visual stimulus is fixed (0.2 seconds), the onset time of the auditory stimulus varies systematically (from 0.25 to 0.45 seconds). The end behavior, "sameness call", of our model for this task has already been validated by experimental results in section 5-1, the blue lines (spatial disparity around 7^{o}) in figures 2-A and 2-C, and we want to show the internal dynamics here. Figure 4B shows the similarity measure, represented in the multisensory dimension of the decision variable, for five temporal disparities. Figure 4C shows the sameness calls, represented in the multisensory dimension of the decision result. For a fixed spatial structure, the similarity measure decreases when the temporal disparity increases. There is a point, around 0.1 seconds of temporal disparity for this case, that the decision about the uniqueness of the cause changes. Based on the mechanism proposed in our model, the change in the sameness call occurs when the spatiotemporal similarity between the stimuli falls below threshold which leads to the more reliable of the unisensory plans to win (not shown here). These simulations show how the temporal evolution of the internal system is influenced when the temporal disparity between cross-modal stimuli varies, sometimes leading to a change in decision through time (dt = 0.1(s) or 0.15(s) here).

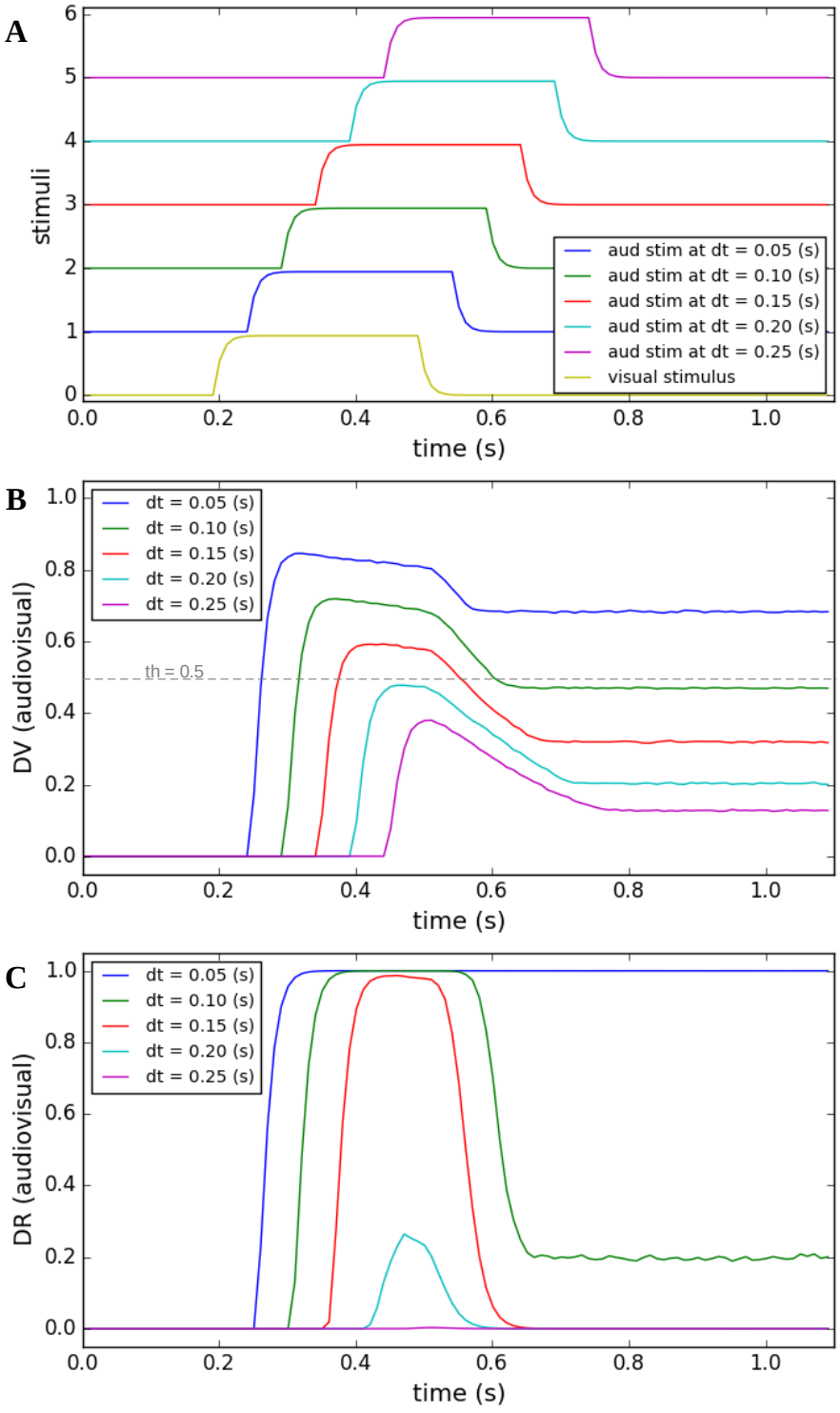

Figure 4: **Effect of temporal disparity of cross-modal stimuli on target selection.** Five different conditions are considered (illustrated by color coding). The spatial and temporal features of the visual stimulus and the spatial features of the auditory stimulus are fixed for all conditions. The onset time of the auditory stimulus varies from 0.25 to 0.45 sec. **A)** The temporal profile of the visual stimulus (lower curve, fixed) and the auditory stimulus (5 upper curves, changing). **B)** The multisensory component of the decision variable is shown for different conditions as a function of time. It changes for different conditions based on the temporal disparity of the stimuli in each condition. The unisensory components (not shown) don't change for different conditions. **C)** The multisensory component of the decision result is shown for different conditions. It is "1" (single source) for smaller temporal disparities and changes to "0" (multiple sources) when the temporal disparity exceeds the threshold (shown as a dashed line in B).

2.5.4 Effect of Stimulus Reliability

For the cases in which there is a large spatiotemporal misalignment between the two stimuli, human subjects often infer that two separate sources exist (Chen and Vroomen, 2013, Ursino et al., 2014) and plan a gaze-shift toward the more salient of the two separate signals. In figure 5 we show the performance of our model under a task in which the visual and auditory stimuli are far from each other in space. The spatiotemporal structure is fixed, and the reliability of the visual stimulus (0.5) is also not changing. The variable factor is the reliability of the auditory stimulus which is changing from unreliable (0.2) to highly reliable (0.8) in four conditions (Figure 5A). Figure 5B shows how the decision variable changes through time for the four conditions. The multisensory (crosses) and visual dimensions (dashed lines) of the decision variable are the same for all conditions, but the auditory dimension is different under each condition because the reliability of auditory stimulus changes. Figure 5C shows result of the auditory plan winning, represented in the auditory dimension of the decision result, for each condition. At the time 0.4 (s) the multisensory component of the decision variable drops below the threshold (figure 5-B), the multisensory component of the decision result changes from zero to one, the unisensory component of the decision result (corresponding to the more reliable stimulus) changes from one to zero, and two separate sources are recognized. When the reliability of visual stimulus is higher than the auditory stimulus the visual plan wins, and if it is lower the auditory plan wins. These simulations show how the temporal evolution of the internal system is influenced when the reliabilities of stimuli vary, leading to selection of the more reliable stimulus as the goal.

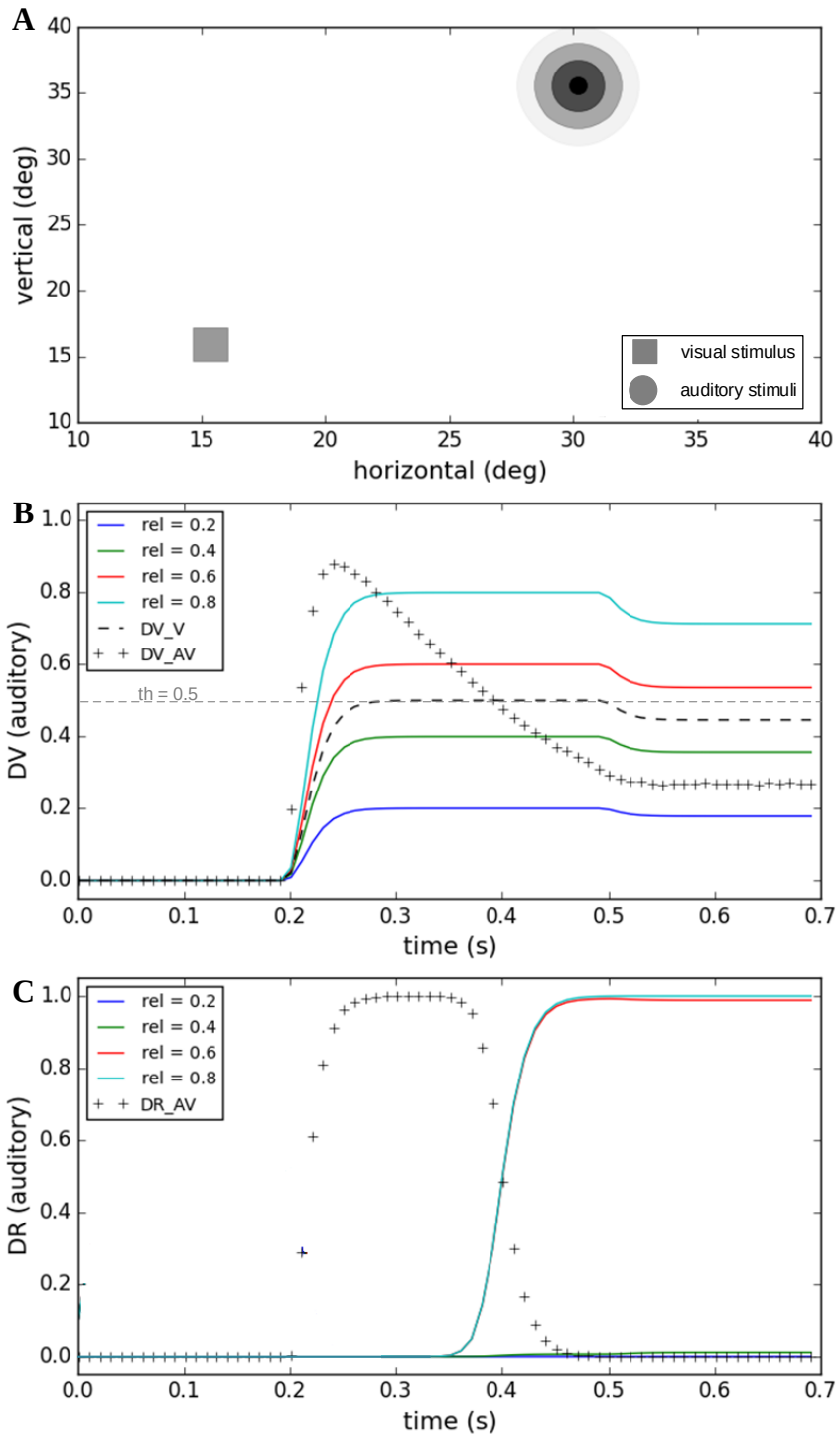

Figure 5: **Effect of the reliability of the unimodal stimuli on target selection.** Four different conditions are considered (illustrated by color coding) The spatiotemporal features of both stimuli are fixed and are chosen such that the similarity measure is always small enough that separate causes are distinguished in all conditions. **A)** The visual stimulus with fixed reliability is shown by a square. The auditory stimulus with varying reliability is illustrated by concentric circles of different levels of blur. **B)** The decision variable is shown for different conditions as a function of time. The visual (thick dashed line) and multisensory components (line of crosses) are the same for all conditions. The auditory component (solid colored lines) varies between different conditions based on the reliability of the auditory stimuli, as shown in A. **C)** The decision result for the auditory component is shown for different conditions as a function of time. The multisensory component (line of crosses) is the same for all conditions. The auditory component is unity when the reliability of the auditory stimulus is higher than the visual stimulus and changes to zero when the auditory stimulus is more reliable than the visual stimulus. The visual component of decision changes in the opposite way.

2.5.5 Effect of Evidence Accumulation

Accumulation of evidence may lead the decision to lean towards an alternative category other than the currently preferred category (Gold and Shadlen, 2007). This has been observed in many oculomotor tasks, for instance, in "anti-saccade" task where the subjects, by default, would plan a saccade towards the presented target, unless some instructive cue commands them to plan a saccade in the mirror opposite direction to the target, in contrast to the default (Everling and Fischer, 1998, Munoz and Everling, 2004). Another example is the "saccade countermanding" task where the subject, by default, has to make a saccade toward the visual target, unless some cue instructs it to stop the motor plan and keep fixating (Hanes and Schall, 1995, Schall et al., 2000). In our case, when stimuli from multiple modalities are presented, we postulate that the default is to assume a common cause for them. This default can be changed to another decision, i.e. separate causes, by accumulation of evidence over time. This concept has been materialized in our model by the development of the similarity measure and its effect on the decision result. We illustrate this concept in two tasks shown in the left and right columns of figure 6.

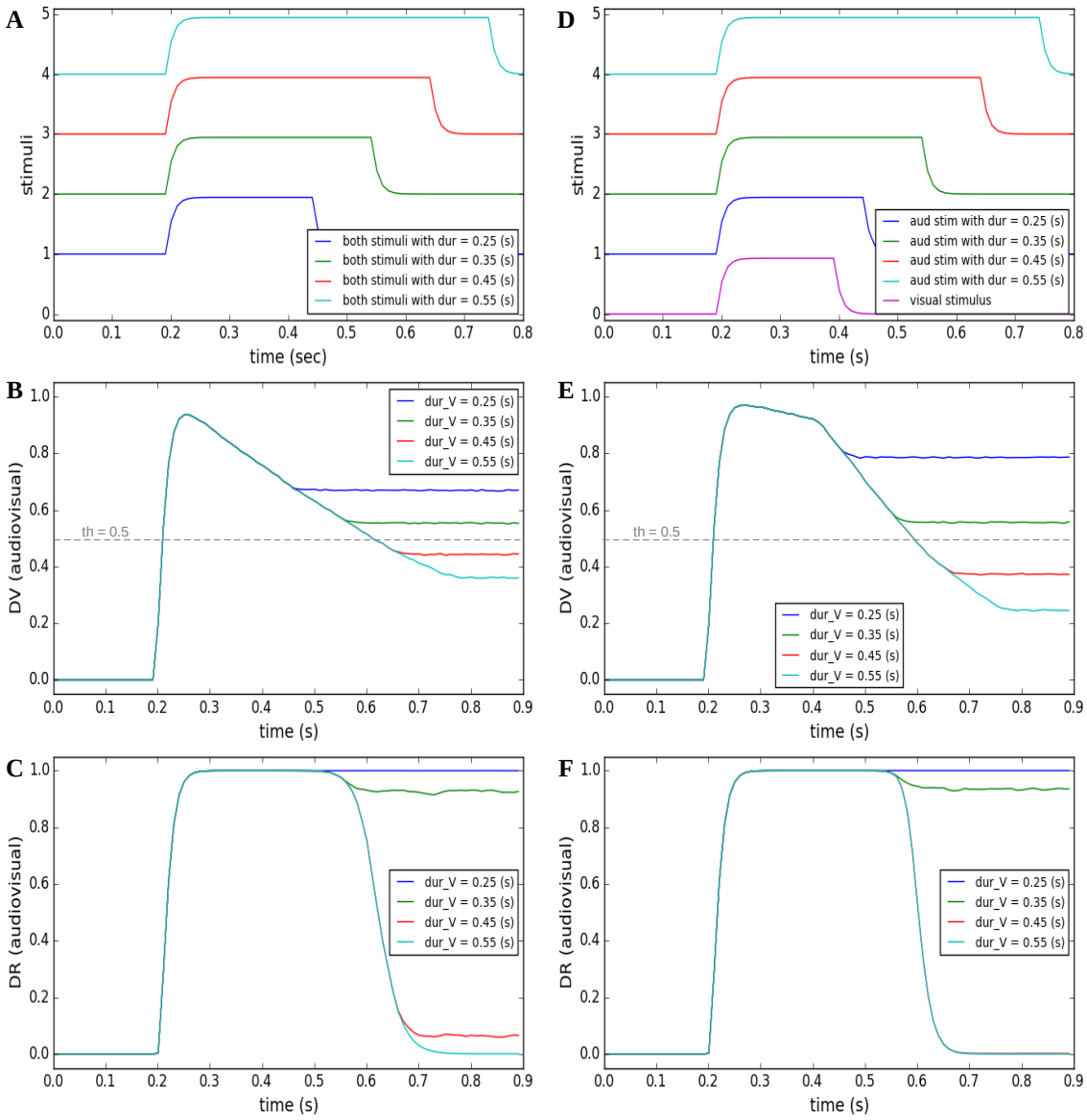

Figure 6: **Effect of accumulation of evidence about cross-modal stimuli on changing target selection decision.** In each of the columns (A, B, C and D, E, F) four different conditions are considered (illustrated by color coding). In the left column, the temporal features of the two stimuli are exactly the same. The stimuli are presented at a fixed, small spatial distance from each other in all conditions. Only the duration of presentation of the stimuli varies for the different conditions (from 0.25 to 0.55 sec). In the right column, the spatial and temporal features of the visual stimulus are fixed (purple curve in D). The two stimuli have a same onset time (0.2 sec) and are presented at a fixed distance from each other, in all conditions. However, the duration of presentation of the auditory stimulus changes from 0.25 to 0.55 (s) (curves 1-4 in D). **A, D)** temporal profiles of the stimuli. **B, E)** The multisensory component of the decision variable is shown for different conditions as a function of time. It changes for different conditions. The unisensory components do not change for different conditions (not shown). The threshold value is shown as a horizontal dashed line. **C, F)** The multisensory component of the decision result is shown for the different conditions. It is initially unity (common cause) first when the stimuli appear. However, it may change to zero (separate causes) if and when enough evidence has accumulated to support the existence of two separate causes.

The left column shows the model's predictions for a case where two stimuli are presented at fixed positions close to each other. As illustrated in figure 6A, the duration of time that the stimuli are present is varied (from 0.25 seconds to 0.55 seconds). Figure 6B shows the similarity measure, represented in the multisensory dimension of the decision variable (dv_{av}), and figure 6C shows the sameness call, represented in the multisensory dimension of the decision result (dr_{av}), developing across time. When the two stimuli are presented briefly and at the same time, they are perceived as belonging to a common source even if they are not presented at exactly the same position in space. But for the same spatial configuration, if the duration of stimulus presentation increases, the similarity measure decreases. There is a point, around 0.4 seconds of presentation duration for this case, that the decision about the uniqueness of the cause changes.

The right column shows the model's prediction for a case where one stimulus appears briefly but the other stimulus might stay on for a longer time. The auditory and visual stimuli, presented at fixed positions very close to each other, have the same onset time (0.2 seconds) but the auditory stimulus is on from 0.05 to 0.35 seconds longer than the visual stimulus (which has a duration of 0.2 seconds) (figure 6D). Figure 6E shows the similarity measure, represented in the multisensory dimension of the decision variable, and figure 6F shows the sameness call, represented in the multisensory dimension of the decision result, developing over time. By extending the presentation duration of one stimulus, while the other is presented only briefly, the similarity measure decreases. Therefore, the sameness decision which was for a common source for shorter durations changes to being for separate sources for longer durations. These examples show that the default decision (that stimuli arise from a common cause) can be altered over a period of time during which evidence accumulates indicating (perhaps) that they are in fact separate. The duration over which evidence needs to accumulate may correspond to the temporal binding window.

2.6 Discussion

In summary, we have proposed a computational model of the cognitive internal system underlying causal inference in spatial localization of cross-modal stimuli. The emerging output of this internal system (report of sameness), not itself, is measurable by psychophysicists. We first showed that our model can replicate the behavioral reports of

the perception of a common cause measurable in psychophysical experiments. Having verified the model, we then moved on to illustrate the dynamics of the decision variable and decision result when spatial and temporal features of the stimuli were changing, like the existing tasks. We then showed the system dynamics for novel situations were separate causes would be inferred or when the decision would change from common to separate sources through evidence accumulation. These dynamic simulations may be tested by new experiments that force the subject's report at specific times and see if the decision changes based on the timing of this forced decision.

Importantly, this new model incorporates several novel features that we expect to be valuable for understanding multisensory integration in the real brain. Based on the ability of our model to replicate known behavioral results (References), and contingent on the further verification of our model's new predictions, we propose that 1) the brain's distributed working memory is multisensory and should retain and process the sensory information to perform this task. 2) Separate computational units are required for representing alternative plans (probably in the cortex) whose selective inhibition (perhaps through basal ganglia connections to cortex) implements the result of the decision. 3) A central decision-making unit should exist capable of applying decision rules, and choosing between multiple causal scenarios based on sensory evidence. 4) Our spatiotemporal similarity measure, capturing how similar the spatial and temporal features of the stimuli are, is the criterion for inferring a common cause. In short, we suggest that the real brain incorporates similar features as our model at the computational level. Further, the current computational-level model is constructed in such a way as to provide a potential formal framework for models that generate physiological predictions at the level of single units and networks.

Finally, the model framework that we have proposed here (simulating causal inference from one visual and one auditory stimulus) has the potential to generalize to a number of other, more complex situations where working memory is a limiting factor. For example: 1) one can tackle target selection between more than two stimuli (Schall and Hanes, 1993, Hill and Miller, 2010) by enhancing the capacity of our short-term memory, increasing the number of possible plan representations and the dimensions of the decision variable, and

defining a multi-dimensional distance variable. 2) One can address causal inference and integration for other modality combinations like visual / tactile and auditory / tactile (Menning et al., 2005, Katus et al., 2015). 3) One can address a situation where a subject has a prior expectation of where the target would appear (Van Wanrooij et al., 2010). When the target is presented one has a causal inference problem to solve, which is whether or not the presented and expected signals are the same, and whether or not to integrate the internal and sensory representations. 4) One can extend the features of the stimuli to include semantic or emotional values (Robertson, 2003). This requires expansion of our concept of similarity to include the more cognitive and linguistic aspects assigned to the stimuli.

3 Reaction Time Variability of Multimodal Gaze-Shifts: A Computational Study in a Decision Making Framework

Mehdi Daemi[1, 2, 3, 4], J. Douglas Crawford[1, 2, 3, 4, 5, 6 †]

1 Department of Biology and Neuroscience Graduate Diploma, York University, Toronto, ON, Canada

2 Centre for Vision Research, York University, Toronto, ON, Canada

3 Canadian Action and Perception Network

4 Department of Psychology, York University, Toronto, ON, Canada

5 School of Kinesiology and Health Sciences, York University, Toronto, ON, Canada

6 NSERC CREATE Brain in Action Program, York University, Toronto, ON, Canada

Submitted

† **Correspondence:**
Dr. J. Douglas Crawford,
Center for Vision Research
Room 0009, Lassonde Bldg.
York University
4700 Keele Street
Toronto, Ontario, Canada, M3J 1P3
jdc@yorku.ca

3.1 Abstract

When goal directed movements are aimed toward multimodal stimuli, cognitive processes during the delay period can influence action planning in both the spatial and temporal domains. In our previous paper (Daemi et al. 2016) we modeled causal inference, based on the spatiotemporal features of multimodal (visual and auditory) stimuli, in order to determine the *spatial* location of goal for saccade. Here, we extend this framework in the *temporal* domain, proposing that "confidence" on selecting a winning plan relative to other alternatives should influence the timing of execution of the winning action plan.

To model these concepts we build upon the evidence-accumulation decision-making framework we previously used to solve the causal inference problem for saccades (Daemi et al. 2016). Once a winning motor plan has been chosen based on causal inference, an instantaneous measure of confidence is computed based on the relative saliency of the winning motor plan compared to the alternate plans. A winning plan is only initiated when enough evidence is accumulated in its favor. This is realized by introducing an accumulative measure of confidence which integrates the instantaneous measure through time. A threshold is then set on the accumulative confidence and a GO command is released whenever it reaches the threshold.

Using this model, we produced simulations that replicate and explain several experimental multisensory observations, including: 1) Lower reaction time for unimodal targets of higher reliability due to more confidence on a unimodal plan. 2) Higher reaction time for multi-modal targets of higher reliability due to less confidence on a unique cause. 3) Higher reaction time for more spatially distant multi-modal targets due to less confidence on a same origin. 4) Higher reaction time for more temporally distant multi-modal targets due to less confidence on a unique cause. Thus, our model provides a unified viewpoint to explain, for the first time, the effects of both spatial and temporal factors on reaction time variability, and assign each of these effects to a unique cognitive function upstream from sensorimotor transformations.

3.2 Introduction

Reaction time (RT) is a measure of speed with which a subject responds to the stimuli within the context of a task. RT has been used to investigate hypotheses about the mental and motor processes to implement different tasks (Sternberg, 1969). In multisensory integration (MSI) research specifically, RT has been used to assess how combining multimodal stimuli with various intensities affect task implementation and response generation (Hershenson, 1962, Rubinstein, 1964). Here, we contemplate the possibility of a unified mechanism which can explain the effects of spatial, temporal, and intensity features of cross-modal stimulation on RT.

It is well known that bimodal stimuli, e.g. visual and auditory, affect the reaction times of goal-directed saccadic eye movements. In particular, when the two stimuli are aligned in space and time, a considerable reduction of the saccade RT is typically observed relative to visual stimulus alone or to auditory stimulus alone. Conversely, RT increases more slowly or even decreases when the stimuli are presented farther from each other or when the delay between them gets larger (Frens et al., 1995, Corneil et al., 2002, Diederich and Colonius, 2004, Navarra et al., 2005, Diederich and Colonius, 2008a, b, Navarra et al., 2009, Van Wanrooij et al., 2010).

Through the years, there have been various attempts to model the variability of RT in multisensory tasks. They have mostly focused on the effect of temporal configuration of the cross-modal stimuli on the RT. The first group of models is referred to as "separate activation" or "race" models. They assume parallel and completely separate channels of sensory processing for stimuli from different modalities. Each channel builds up some independent activation. Response is triggered by the channel which reaches some threshold level first. Average RT to multisensory stimuli is lower than unimodal stimuli because the average of winner's processing time is smaller than average processing time in each single channel (statistical facilitation). Independent Gaussian distributions (Raab, 1962) and experimentally observed distributions (Gielen et al., 1983) were used as unimodal distributions to estimate the minimum distribution in the bimodal conditions. Nevertheless, statistical facilitation couldn't account for facilitation in data (Diederich and Colonius, 2008a).

The second group of models is called “coactivation” models. They assume that activation raised in different sensory channels by presenting multimodal stimuli is combined to satisfy a single criterion for response initiation. The discrete realization of this idea gave rise to the so-called superposition models while its continuous realization brought about multichannel diffusion models (Schwarz, 1989, Diederich, 1992). In all such models, the stimulus intensity is represented by some internal indicator (“counter” for superposition models and “drift” for diffusion models). For multimodal stimuli, these internal variables from multiple sensory channels are added together during some peripheral stage of processing. This leads to faster reaching some threshold (fixed number of counts for superposition models and threshold limits for diffusion models) and lower RT.

Previous models could not account for distinguishing a target modality from a nontarget modality in experiments like the focused attention paradigm (Amlot et al., 2003, Diederich and Colonius, 2007). To consider such effects, time-window-of-integration (TWIN) models combine basic ideas of the previous groups of models (Colonius and Diederich, 2004). They consider two stages of processing. The first stage consists of separate and parallel processing in unisensory pathways. The second stage comprises the combination of the unisensory activations and response initiation. Second stage occurs only if the peripheral processes of the first stage all terminate within a given time interval (Colonius and Diederich, 2010). Such two-stage models support the idea that the race between the sensory channels takes place upstream from the SC, and that the SC itself is part of the second stage (Sparks and Mays, 1990a).

None of these models, briefly described above, account for effects of spatial disparity of the cross-modal stimuli on the reaction time. These models also ignore the internal perception of the subjects, namely whether they perceive the multimodal stimuli as belonging to a unique event to separate events. Here we propose a model which explains the variability of saccadic reaction time as a function of both temporal and spatial configurations of the stimuli. We build this model upon a previous model of causal inference spatial localization in multi-modal situations (Daemi et al. 2016). A decision-making circuitry was proposed, where different plans represented different possibilities for the inferred cause, namely same or separate sources. A spatiotemporal similarity measure

was proposed, and was compared to the reliabilities of the unimodal stimuli to make the decision on the causal structure. Here we extend that framework to model how the timing of an action based the inference is determined. The saccadic reaction time is proposed to depend on a measure of accumulated confidence on the decision made about the causal structure. This expanded framework can explain variability of the reaction time as functions of 1) spatial configuration of the stimuli 2) temporal configuration of stimuli 3) reliabilities of the stimuli. This means we can explain the interesting patterns of variability in reaction times of gaze-shifts towards cross-modal stimuli, which could not be explained based on psychophysical models of sensory-driven reactions. We do so by proposing an internal model and assigning a cognitive significance, namely accumulative confidence, to the factor governing the reaction time.

3.3 Model Overview

Generally, in terms of action initiation, two types of tasks are possible: 1) a forced-reaction-time (forced-RT) task, and 2) a choice-reaction-time (choice-RT) task. In the forced-RT task the time for onset of the action is a requirement of the task and is forced by a higher-order, top-down command. In the choice-RT task the subject is free to start the action as soon as it is ready. Here, we want to build a model of gaze-shift initiation, in multi-modal situations, for a choice-RT task. We propose that the readiness for initiating a winning action plan is measured by the confidence on the decision that determined that plan is winning.

Consider a situation where visual and / or auditory stimuli, with possible spatial and temporal disparity and different intensities are presented to a subject who is instructed to make a gaze-shift to the most reliable of the targets. We previously proposed a model of how subjects may infer the cause of the stimuli, a common source or separate sources, by introducing a spatiotemporal measure of similarity (Daemi et al. 2016). Here we extend that model to account for the variability of multi-modal reaction times. Figure 1 illustrates this extended, unifying framework, containing three color-coded sections, explained in the following paragraphs.

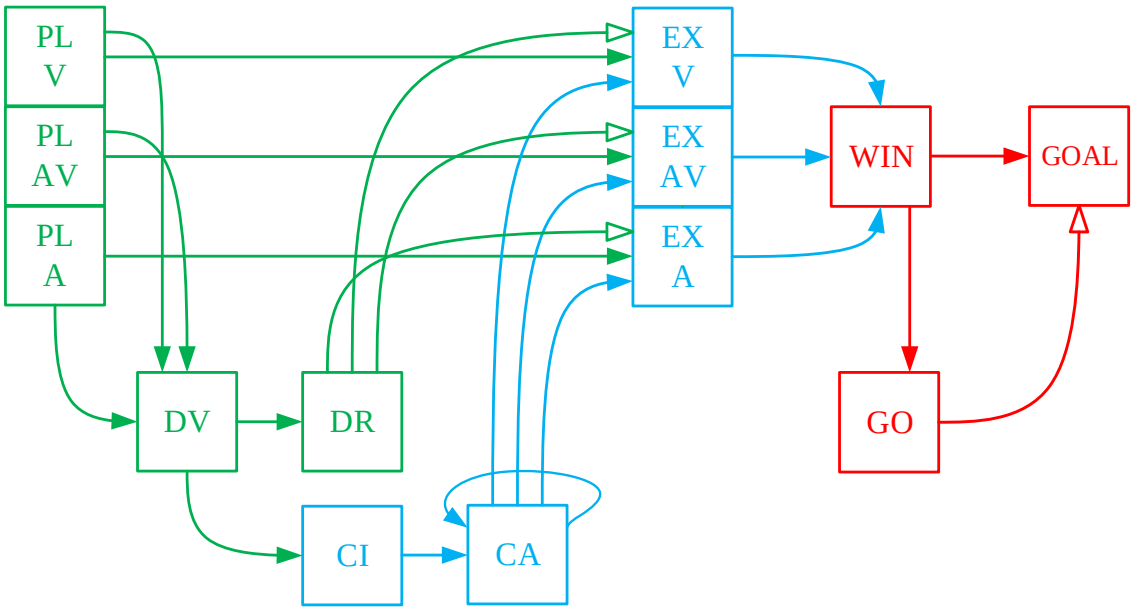

Figure 1: **The model of reaction time variability of gaze-shifts towards cross-modal stimuli.** Upstream of this model (not shown), the spatiotemporal similarity between stimuli was measured and the multimodal signals were integrated in working memory. The three possible plans were constructed in units PL_V, PL_AV, PL_A as saliency spatial maps. The decision variable is constructed in the unit DV by sending the saliency of the three plans as their bids. The decision is made in the decision result DR, by materializing the concept that if the similarity measure is bigger than a threshold, then the multisensory plan wins and if not, the unisensory plan with higher reliability wins. The instantaneous, CI, and accumulative, CA, measures of confidence on the decision are calculated from the decision variable. The plan representations in the execution layer, three units EX_V, EX_AV, EX_A, are constructed as confidence maps buy communicating the spatial information from plan layer and the corresponding accumulative confidence value. The decision result is realized by selective inhibition of the plan units in execution layer (not shown). The confidence map of the winning plan is then sent to the unit WIN. The confidence value of the winning plan is sent to GO. The spatial information of the WIN is sent to GOAL. GOAL is under constant inhibition of GO. When the confidence reaches a threshold, GOAL is disinhibited.

The green parts of the model include the decision making circuitry underlying causal inference. This is where the three alternative plans, realizing the three possible solutions to the causal inference problem, are represented. These include a unisensory plan ($\overrightarrow{PL\_V}$) that manifests separate sources and gaze-shift to the visual signal, another unisensory plan ($\overrightarrow{PL\_A}$) that manifests separate sources and gaze-shift to the auditory signal, and a multisensory plan ($\overrightarrow{PL\_AV}$) that manifests a common source and gaze-shift to the weighted average of their spatial positions. Each of these plans include a saliency component, and the decision on the causal structure is made by systematically comparing these saliencies. The saliencies are of the unisensory plans are the reliability of the unimodal stimuli (that can be simplified to their intensities). The saliency of the multisensory plan is a measure of spatiotemporal similarity between the cross-modal stimuli. For making the decision, the saliencies of the alternative plans are sent to construct a decision variable ($\overrightarrow{DV}$). A decision rule is then applied on the decision variable in its transformation to the decision result ($\overrightarrow{DR}$). The decision rule realizes a specific comparison between the plan saliencies: the multisensory plan is chosen if its saliency is greater than a threshold, and the more reliable of the unisensory plans is chosen if the similarity measure is smaller than the threshold.

The blue parts of the model are involved in calculating *when* to send the winning plan for execution and initiation of the action. An instantaneous measure of confidence ($\overrightarrow{CI}$) is first constructed by transforming the decision variable, realizing the idea of immediate confidence on a plan to guide action if it is winning the competition among alternative plans. So, at any time, $\overrightarrow{CI}$'s component corresponding to a plan is zero if that plan is not winning and it is equal to the momentary confidence on the decision that the plan is winning, if it is winning. As the decision result could be changing by accumulation of evidence, a $\overrightarrow{CI}$ component could be changing between zero and a positive value through time. Then, an accumulative measure of confidence ($\overrightarrow{CA}$) is constructed by integration, through time, of the instantaneous measure of confidence. This measure reflects accumulation of evidence, through time, against or in favor of a plan to guide the action. Each component of $\overrightarrow{CA}$ indicates the amount of confidence on implementing a plan, accumulated through time.

We have salience maps of space in the plan layer ($\overrightarrow{PL\_V}$, $\overrightarrow{PL\_V}$, $\overrightarrow{PL\_V}$), which represented alternative causal structure. In order to implement the decision on the preferred causal structure, another layer of plan representations is constructed in an execution layer. This involves three confidence maps of space, which receive their spatial components from the corresponding representations in the plan layer, and their confidence components from corresponding components of the accumulative measure of confidence. They include a unisensory map ($\overrightarrow{EX\_V}$) that manifests the victory of the visual plan, another unisensory map ($\overrightarrow{EX\_A}$) that manifests the victory of the auditory plan, and a multisensory map ($\overrightarrow{EX\_AV}$) that manifests the victory of the multisensory plan. These plan representations are selectively inhibited by the decision result to implement the selection of the inferred causal structure.

The red parts of the model are involved in implementing the timing of action initiation. In previous parts, a spatial plan was chosen to guide the action, and the confidence on the decision to choose that plan was calculated. All this information is reflected in the one confidence map of space, in the execution layer, which was disinhibited by the decision result. That winning plan is now communicated to a computational unit called $\overrightarrow{WIN}$. The spatial plan, only, is then communicated from $\overrightarrow{WIN}$ to another computational unit called $\overrightarrow{GOAL}$. However, $\overrightarrow{GOAL}$ is constantly inhibited by a GO command, by default, when we are not confident enough to execute a plan. The GO command is constructed by applying a threshold function on the confidence measure in $\overrightarrow{WIN}$. The result of this function is that the GO signal is 'zero' if the $\overrightarrow{WIN}$'s confidence measure is smaller than a threshold and it becomes 'one' if the confidence rises above the threshold. $\overrightarrow{GOAL}$ is inhibited by GO if GO is 'zero'. However, if GO becomes 'one' the $\overrightarrow{GOAL}$ is disinhibited and the spatial plan of WIN is allowed to be communicated into $\overrightarrow{GOAL}$. $\overrightarrow{GOAL}$ then sends the winning plan to the machinery involved in eye-head coordination and implementing the gaze-shift.

The general outline of the model is inspired by known properties of the decision making, action selection, and gaze control systems in the brain. The model's representations of alternative plans involved in decision making, i.e. the plan (PL) and execution (EX) layers, are inspired by such neural codes in frontal cortex (Jones et al., 1977, Canteras et al., 1990,

Berendse et al., 1992, Yeterian and Pandya, 1994, Levesque et al., 1996). A central arbitrating system is thought to receive bids, plan saliencies as in our case, from alternative plans for further processing (Redgrave et al., 1999). This information processing, e.g. in the telencephalic decision centers, underlies constructing a decision variable, implementing a decision rule, computing a decision result, and calculating the confidence on the decision. The units $\overrightarrow{DV}$, $\overrightarrow{DR}$, $\overrightarrow{CI}$, $\overrightarrow{CA}$, and $\overrightarrow{GO}$, the internal connections between them and their projections from plan representations have been inspired by the known physiology of this central decision making system in the brain (Gold and Shadlen, 2007, Cisek and Kalaska, 2010). The basal ganglia are thought to receive the result of the decision from cortex (Beiser and Houk, 1998, Koos and Tepper, 1999, Gernert et al., 2000) and implement it through selective disinhibition of cortical and sub-cortical representations. This is realize in our model, through a multiplicative effect on plan representations, in two occasions: 1) selective inhibition of the execution layer EX by the decision result $\overrightarrow{DR}$, 2) selective inhibition of the goal representation $\overrightarrow{GOAL}$ by the 'go' command GO. We assume the winning plan in EX is sent to the gaze control system to plan a gaze-shift (while it could possibly be sent to other motor circuitries to plan a reach or grasp, for example). So, our final spatial maps are specifically inspired by the saccade-related neural populations in the superior colliculus (Munoz and Wurtz, 1995b, a): 1) confidence map in $\overrightarrow{WIN}$ is thought to be implementing the function of the buildup neurons, 2) motor map in $\overrightarrow{GOAL}$ is inspired by the physiology of the burst neurons. The final winning plan $\overrightarrow{GOAL}$ is then assumed to be sent to the brainstem (Sparks, 2002, Girard and Berthoz, 2005) to drive the eye-head coordination system (Klier et al., 2003, Daemi and Crawford, 2015) to reorient the line of sight to the selected target.

3.4 Mathematical Formulation

3.4.1 Method

The proposed model incorporates a concept "confidence" on the selected plan within a decision making framework as the underlying factor which determines the timing of its execution (see Model Overview), and we show that this concept can explain the complex variability of reaction times in cross-modal configurations (see Results). We are linking the time dimension of action planning to the dynamics of the evidence-based decision

making, a high-level cognitive process. To model this, we need to move beyond 1) the classic cognitive architectures that neglect the time dimension of inferential and logical transformations (Newell and Simon, 1972, Anderson, 1983), and 2) traditional approaches, like classic-control theory and cybernetics, which constrain goal-directed motor planning with real time constraints of environmental interactions, but ignore the high-level cognitions (van Gelder, 1998). A more general framework realizes cognitive processes within the time constraints of interacting with and surviving in a dynamically changing environment, just like the brain.

In this model, an approach that considers "perception-action" and "high-level cognition" in a unified framework (Eliasmith, 2013) has been adopted. Inspired by the brain neurophysiology (Fuster, 2005), models within this more general framework are implemented in distributed networks of parallel processing units. This characterizes sensorimotor and cognitive transformations by functions of both the internal state variables and the time, realized in connections between the units. Routing of information between the units (attentional control), through time, is flexibly controlled. Even though we do not address how the model could be realized in a neural architecture in the current study, most units and their connections in the model were developed based on the known neurophysiology, and can be neurally implemented by the Neural Engineering Framework, a recent method that unifies the symbolic, connectionist, and dynamicist viewpoints (Eliasmith and Anderson, 2003, Eliasmith et al., 2012). Modelling an adaptive, robust biological system which can behave and survive in an uncertain environment justifies the relatively high number of variables in such models. Taking this approach, we modeled causal inference, in an evidence-based decision making circuitry, as a process evolving through time which can help us dynamically interact with the environment, executing actions in proper times.

3.4.2 Decision Making

Multimodal signals are detected from the environment and encoded in early sensory areas whose dynamics reflect the temporal aspect of the stimuli presentation. These sensory signals are then communicated to the working memory to be retained and further processed. Spatiotemporal similarity measure, our criterion to infer the origin of the stimuli, is

calculated from the sustained signals in the working memory. The plan representations in PL are then constructed to manifest the possible causal structures. The two unimodal plans represent the spatial position of the unimodal stimuli along with their reliabilities as the plan saliencies (sal_{plv} and sal_{pla}). The multisensory plan represents the weighted average of the unimodal position signals along with the spatiotemporal similarity measure as plan saliency (sal_{plav}). The decision variable is constructed by the saliencies of the plans:

$$\overrightarrow{DV}(t) = \begin{bmatrix} dv_v \\ dv_a \\ dv_{av} \end{bmatrix} = \begin{bmatrix} sal_{plv} \\ sal_{pla} \\ sal_{plav} \end{bmatrix} \quad (1)$$

The decision rule is realized by a transformation of the $\overrightarrow{DV}$ resulting in decision result, $\overrightarrow{DR}$:

$$\overrightarrow{DR}(t) = \begin{bmatrix} dr_v \\ dr_a \\ dr_{av} \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{-sl_m(th_{av}-dv_{av})}} \times \frac{1}{1+e^{-sl_u(dv_v-dv_a)}} \\ \frac{1}{1+e^{-sl_{av}(th_{av}-dv_{av})}} \times \frac{1}{1+e^{-sl_u(dv_a-dv_v)}} \\ \frac{1}{1+e^{-sl_{av}(dv_{av}-th_{av})}} \end{bmatrix} \quad (2)$$

th_{av} is a threshold value applied on the similarity measure, determining whether or not a unique object originated the signals. $\overrightarrow{DR}$ controls the implementation of the decision of which plan to drive the gaze-shift. This is applied by selective inhibition of plan representations in execution, cortical layer (EX). Plan representations in EX are governed by these equations:

$$\overrightarrow{EX\_V}(t) = \begin{bmatrix} cnf_{exv} \\ ech_{exv} \\ ecv_{exv} \end{bmatrix} = dr_v \times \begin{bmatrix} ca_v \\ ech_{plv} \\ ecv_{plv} \end{bmatrix} \quad (3)$$

$$\overrightarrow{EX\_A}(t) = \begin{bmatrix} cnf_{exa} \\ ech_{exa} \\ ecv_{exa} \end{bmatrix} = dr_a \times \begin{bmatrix} ca_a \\ ech_{pla} \\ ecv_{pla} \end{bmatrix} \quad (4)$$

$$\overrightarrow{EX\_AV}(t) = \begin{bmatrix} cnf_{exav} \\ ech_{exav} \\ ecv_{exav} \end{bmatrix} = dr_{av} \times \begin{bmatrix} ca_{av} \\ ech_{plav} \\ ecv_{plav} \end{bmatrix} \quad (5)$$

EX plan representations are selectively inhibited to determine the winning plan. This effect is shown by the multiplicative effect of the corresponding $\overrightarrow{DR}$ component. We have actually employed a parallel basal ganglia circuitry for implementation of this selection process. However, we are not including the formulations for the basal ganglia computational units because our main focus is on how the decisions are made but not on how they are implemented. EX plan representations are assumed to be action initiation confidence maps. The first dimension of EX plan representations indicates how confident we are on selecting the corresponding plan if this plan is actually winning. The next section explains how the confidence measures are computed.

3.4.3 Confidence Measure

We intend to propose a criterion of when to initiate a gaze-shift when it is free to implement the decision at any time, i.e. the choice-reaction-time case. Conceptually, it is proposed that this timing is determined by the confidence on the decision. First, an instantaneous measure of confidence ($\overrightarrow{CI}$) is introduced. The confidence on the decision at each time-point is measured by the distance between the bid of the winning plan and the bids of the losing plans. This variable is constructed as a 3-D signal each component of which indicates at any time, if its corresponding plan is winning, how confident the decision that it is winning is.

$$\overrightarrow{CI}(t) = \begin{bmatrix} ci_v \\ ci_a \\ ci_{av} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{(dv_v - dv_{av})^2 + (dv_v - dv_a)^2}}{(1 + e^{-sl_m(th_{av} - dv_{av})}) \times (1 + e^{-sl_u(dv_v - dv_a)})} \\ \frac{\sqrt{(dv_a - dv_{av})^2 + (dv_a - dv_v)^2}}{(1 + e^{-sl_{av}(th_{av} - dv_{av})}) \times (1 + e^{-sl_u(dv_a - dv_v)})} \\ \frac{\sqrt{(dv_{av} - dv_v)^2 + (dv_{av} - dv_a)^2}}{1 + e^{-sl_{av}(dv_{av} - th_{av})}} \end{bmatrix} \quad (6)$$

This confidence grows through time with a slope which is dependent on the value of the instantaneous confidence at any time-point. This concept can be materialized in a structure called ‘accumulative confidence’ by integrating the components of ‘instantaneous confidence’ through time. This is a leaky integrator with a controllable leak. The state space vector of this integrator has four dimensions. The first component is the leak and the other

three components are the accumulative confidence on the corresponding plans to drive the gaze-shift, if it is winning:

$$\overrightarrow{CA}(t) = \begin{bmatrix} lk_{conf} \\ ca_v \\ ca_a \\ ca_{av} \end{bmatrix} \tag{7}$$

The state space equations characterizing this structure are:

$$\begin{bmatrix} \dot{lk}_{conf} \\ \dot{ca}_v \\ \dot{ca}_a \\ \dot{ca}_{av} \end{bmatrix} = \begin{bmatrix} 0\ 0\ 0\ 0 \\ 0\ (1-lk_{mv})\ 0\ 0 \\ 0\ 0\ (1-lk_{mv})\ 0 \\ 0\ 0\ 0\ (1-lk_{mv}) \end{bmatrix} \times \begin{bmatrix} lk_{conf} \\ ca_v \\ ca_a \\ ca_{av} \end{bmatrix} + \begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{bmatrix} \times \begin{bmatrix} ci_v \\ ci_a \\ ci_{av} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times [clear] \tag{8}$$

Conceptually, the subject is to make a gaze-shift when its confidence about its decision reaches a threshold.

3.4.4 GO Command

We propose that all the plan representations in EX, along with their corresponding confidences, converge to the another computational unite that is called $\overrightarrow{WIN}$:

$$\overrightarrow{WIN}(t) = \begin{bmatrix} cnf_{win} \\ ech_{win} \\ ecv_{win} \end{bmatrix} = \begin{bmatrix} cnf_{exv} \\ ech_{exv} \\ ecv_{exv} \end{bmatrix} + \begin{bmatrix} cnf_{exa} \\ ech_{exa} \\ ecv_{exa} \end{bmatrix} + \begin{bmatrix} cnf_{exav} \\ ech_{exav} \\ ecv_{exav} \end{bmatrix} \tag{9}$$

However, because of the selective inhibition applied on EX by $\overrightarrow{DR}$, only one plan, which is winning the decision making process, is feeding $\overrightarrow{WIN}$ at any time. The confidence map of the winning plan in EX is now communicated to $\overrightarrow{WIN}$. The GO command is constructed by transformation of the confidence component of $\overrightarrow{WIN}$:

$$GO(t) = [go] = \frac{1}{1 + e^{-sl_{go}(cnf_{win} - th_{go})}} \tag{10}$$

According to this transformation, as soon as the confidence on the winning plan passes a threshold, the value of the GO signal changes from 0 to 1. The winning plan is communicated from $\overrightarrow{WIN}$ to a final computational unit called $\overrightarrow{GOAL}$.

$$\overrightarrow{GOAL}(t) = \begin{bmatrix} ech_{goal} \\ ecv_{goal} \end{bmatrix} = go \times \begin{bmatrix} ech_{win} \\ ecv_{win} \end{bmatrix} \quad (11)$$

$\overrightarrow{GOAL}$ is under constant inhibition of GO which determines action initiation. When the GO signal has the value zero (the default configuration) the $\overrightarrow{GOAL}$ is inhibited. $\overrightarrow{GOAL}$ gets released out of the GO's inhibition whenever the GO signal becomes one.

3.5 Results

Here the model is used to reconstruct experimental paradigms where saccadic reaction times have been recorded. Hence, the internal mechanisms suggested in the model are verified by reproducing the wide range of the variability of RT during these tasks. We will first investigate the behavior of our model in a simple unisensory task where only one, unisensory stimulus is presented, and a gaze-shift is planned towards it (section 5.1). We then examine the model in cross-modal situations where the spatial and temporal configurations of the presented stimuli vary, similar to previous experimental results, and report their effects on the reaction time (sections 5.2 & 5.3). Next, we use the model to predict the variation of reaction time when the reliability of the cross-modal stimuli change (section 5.4). Finally, we will summarize the results in the last section where we draw the reaction time as functions of temporal disparity, spatial disparity, unimodal stimulus reliability, and cross-modal stimulus reliability (section 5.5).

3.5.1 Unisensory Situation

It has been experimentally observed that reaction time of gaze-shifts towards unimodal stimuli decreases if the intensity of the stimulus increases (Bell et al., 2006). Here we test our model in such a task where only the visual stimulus is present. This is illustrated in Fig. 2, where panel A shows the changing stimulus parameters of the task, whereas panels B-E represent internal states within the model, culminating in the internal GO signal that determines saccade latency.

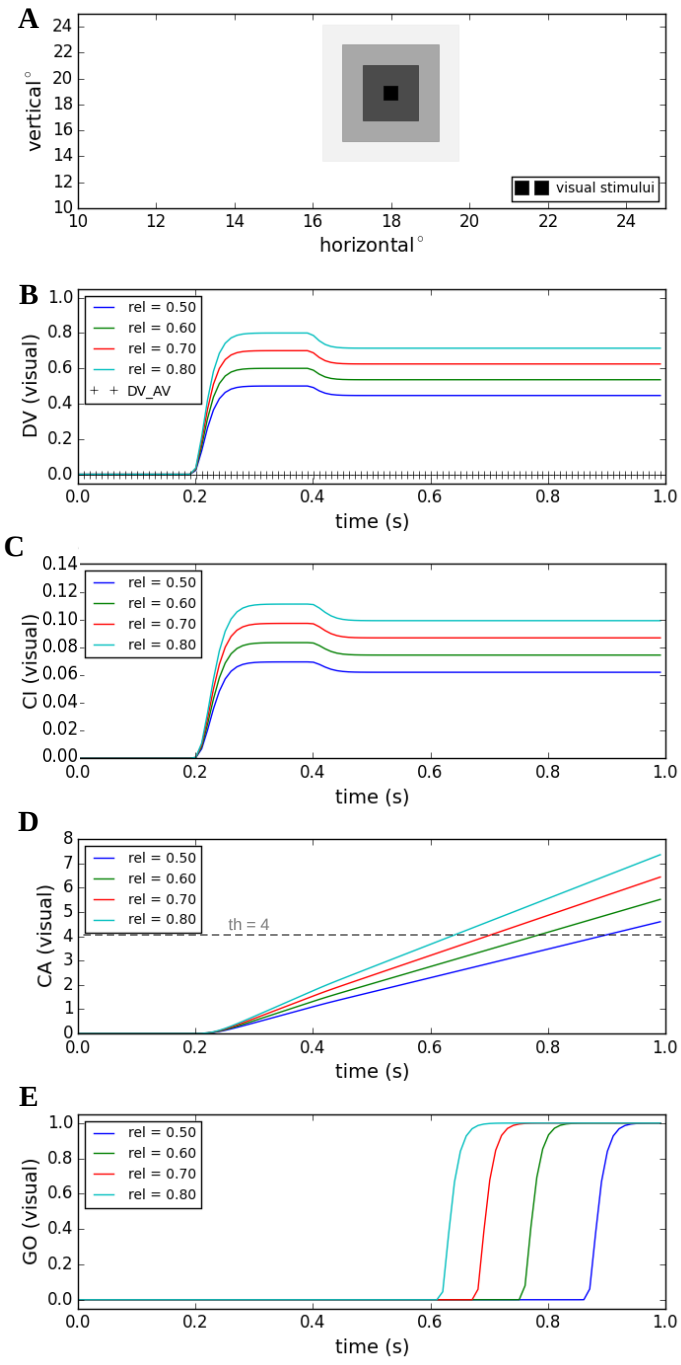

Figure 2: **Effect of target reliability on reaction time when only one stimulus modality is presented.** The auditory target is not presented. The position, onset time, and duration of the visual target is fixed. The reliability of the visual target varies within four different conditions. **A)** The visual stimulus having different reliabilities in different conditions is illustrated by different levels of blurriness. **B)** The decision variable is shown being developed through time. The multisensory and auditory components are zero for all conditions. The visual component is changing between conditions because the reliability (rel) of visual target varies. **C)** The instantaneous confidence on decision is shown through time. Only the visual component has non-zero value. **D)** The accumulative confidence on decision is shown as a function of time. Only the visual component is not zero. **E)** The GO signal is shown for execution of the winning plan. It changes from zero to one whenever the accumulative confidence passes a threshold.

The gray boxes in figure 2-A illustrate visual stimuli with the same spatial and temporal features, but different levels of reliability. The reliability of the visual stimulus changes in four levels, as illustrated in figure 2-A by different levels of blurriness (shades of gray) of the stimulus. Figure 2-B shows the reliability of the visual stimulus in the visual component of the decision variable (color-coded for different conditions). As there exists only one stimulus modality, the similarity measure, represented in the third dimension of the decision variable is always zero (DV_AV in figure 2-B).

The reaction time in such a situation depends on how dominant the visual plan is in the decision variable. This dominance increases when saliency of the visual plan is bigger relative to other plans' saliencies (which are zero) and that happens when the reliability of the visual stimulus increases. This is reflected in higher instantaneous confidence on the decision for higher reliabilities (figure 2-C). Consequently the accumulative confidence reaches its threshold faster relatively (figure 2-D). As a result, the GO command is issued earlier for higher reliabilities of the visual stimulus (figure 2-E).

Thus, our model reproduces the experimental finding that reaction time decreases by increasing the unisensory stimulus reliability (Bell et al., 2006). This is accomplished in the model by implementing the idea that when there is no distracting stimulus, whose relation to the main stimulus should have been inferred, the confidence on a unique stimulus as the target for shifting the attention increases when its reliability increases.

3.5.2 Effect of Spatial Disparity

Reaction time of gaze-shifts towards cross-modal stimuli increases if the spatial distance between the two presented stimuli increases, as experimentally observed (Frens et al., 1995). Here we test our model in such a task. This is illustrated in Fig. 3, where panel A shows the changing stimulus parameters of the task, whereas panels B-E represent internal model variables, culminating in the GO command.

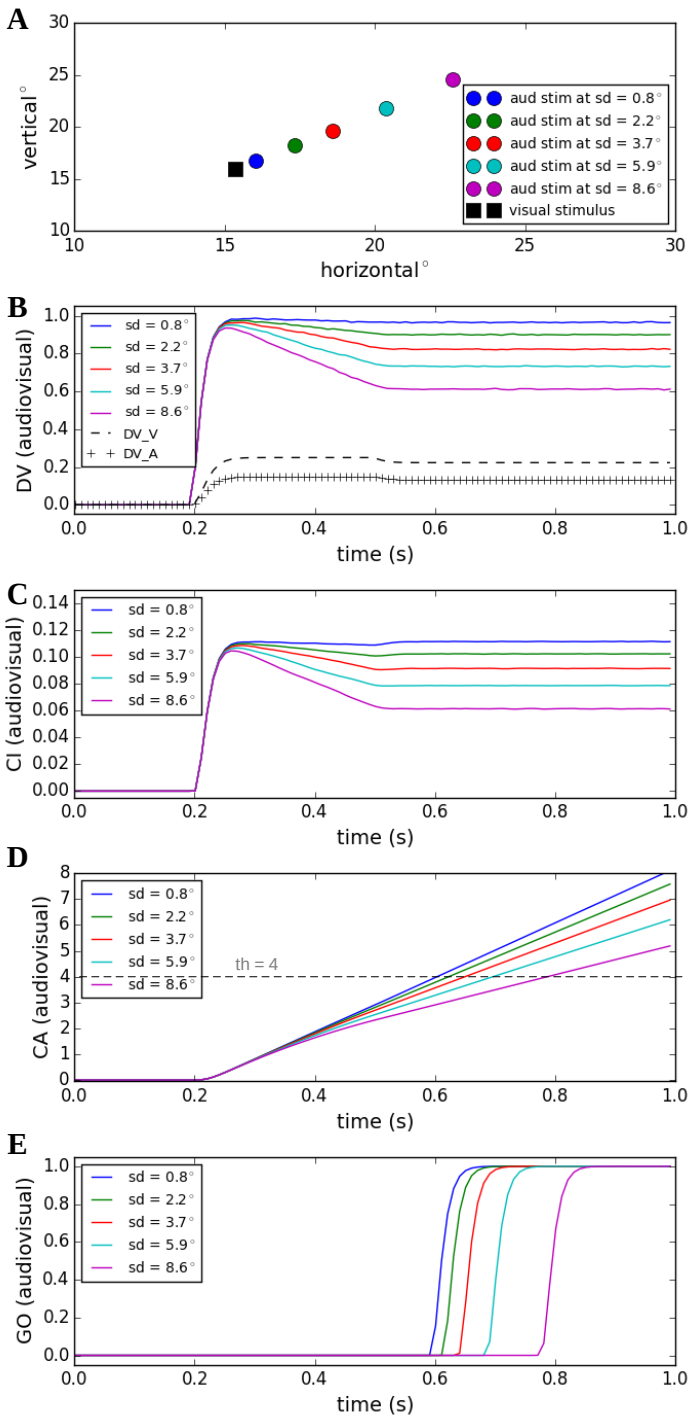

Figure 3: **Effect of spatial disparity of cross-modal stimuli on reaction time.** Five different conditions have been considered (illustrated by color coding). The visual stimulus has a fixed position, onset time, duration, and reliability for all conditions. The onset time, duration and reliability of the auditory target is also fixed. **A)** The spatial position of the auditory target varies, which is reflected in different spatial distances (sd) for different conditions, while the visual target is fixed. **B)** The decision variable is shown being developed through time. The unimodal components are the same for all conditions (dashed and crossed lines). The multisensory component is changing between conditions because the spatial distance varies. **C)** The instantaneous confidence on decision is shown through time. The unimodal components are always zero because the multisensory plan is always winning, because its saliency, i.e. the similarity measure, is greater than the threshold. **D)** The accumulative confidence on decision is shown as a function of time. Only the multisensory component is not zero. **E)** The GO signal is shown for execution of the winning plan. It changes from zero to one whenever the accumulative confidence passes a threshold.

The visual and auditory targets have the same fixed temporal features (onset time 0.2s and duration 0.3s). The position of the visual stimulus is invariable while the position of the auditory stimulus is changing within five conditions, from 0.8^{o} to 8.6^{o} from the visual stimulus (figure 3-A). The reliabilities of the stimuli are invariant as depicted in figure 3-B in the unisensory dimensions of the decision variable (DV_V and DV_A). The similarity measure, coded in the multisensory dimension of the decision variable, is shown in different colors for different spatial distances in figure 3-B. In all conditions the two targets are close enough together that they are perceived to be coming from the same origin.

The reaction time in such a situation depends on how dominant the winning, multisensory plan is in the decision variable. This dominance increases when similarity measure is much greater than the unisensory target reliabilities and that happens by decreasing the spatial disparity (see figure 3-B and compare the saliency of plans). This is reflected in higher instantaneous confidence on the decision for smaller spatial disparities (figure 3-C). Consequently the accumulative confidence reaches its threshold relatively faster. As a result, the GO command is issued earlier for smaller spatial disparities (figure 3-E).

Thus, the model replicates the experimentally found result that the reaction time increases by increasing the spatial disparity between the stimuli (Frens et al., 1995). This is accomplished in the model by implementing the idea that the confidence on the sameness of the origin decreases when the spatial distance between the stimuli increases.

3.5.3 Effect of Temporal Disparity

As experimentally observed (Frens et al., 1995), the reaction time of gaze-shifts towards cross-modal stimuli increases if the temporal disparity between the two presented stimuli increases. Here we test our model in such a task where the visual and auditory stimuli are presented at the same positions and with the same time duration (0.3 (s)) all the time (figure 4). However, while the onset time of the visual target is invariable (0.2 (s)), the onset time of the auditory target changes within five conditions from 0.215 (s) to 0.275 (s) (figure 4-A). The unimodal, stimulus reliabilities are fixed (DV_V and DV_A in figure 4-B). The spatiotemporal similarity measure, coded in the multisensory component of the decision variable, increases for higher temporal disparities as depicted (color coded) in figure 4-B.

In all conditions the two targets are presented close enough together in time, so that they are perceived to be coming from the same origin.

The reaction time in such a situation depends on how dominant the winning, multisensory plan is in the decision variable. This dominance increases when similarity measure is much greater than the unisensory target reliabilities and that happens by decreasing the temporal disparity (figure 4-B). This is reflected in higher instantaneous confidence on the decision for smaller temporal disparities (figure 4-C). Consequently the accumulative confidence reaches its threshold faster relatively (figure 4-D). As a result, the GO command is issued earlier for smaller temporal disparities (figure 4-E).

Thus, the model reproduces the experimental finding that reaction time increases when the temporal disparity between the cross-modal stimuli increases (Frens et al., 1995). The model accomplishes this by implementing the idea that the confidence on the unique-object causal structure decreases when the temporal distance between the stimuli increases.

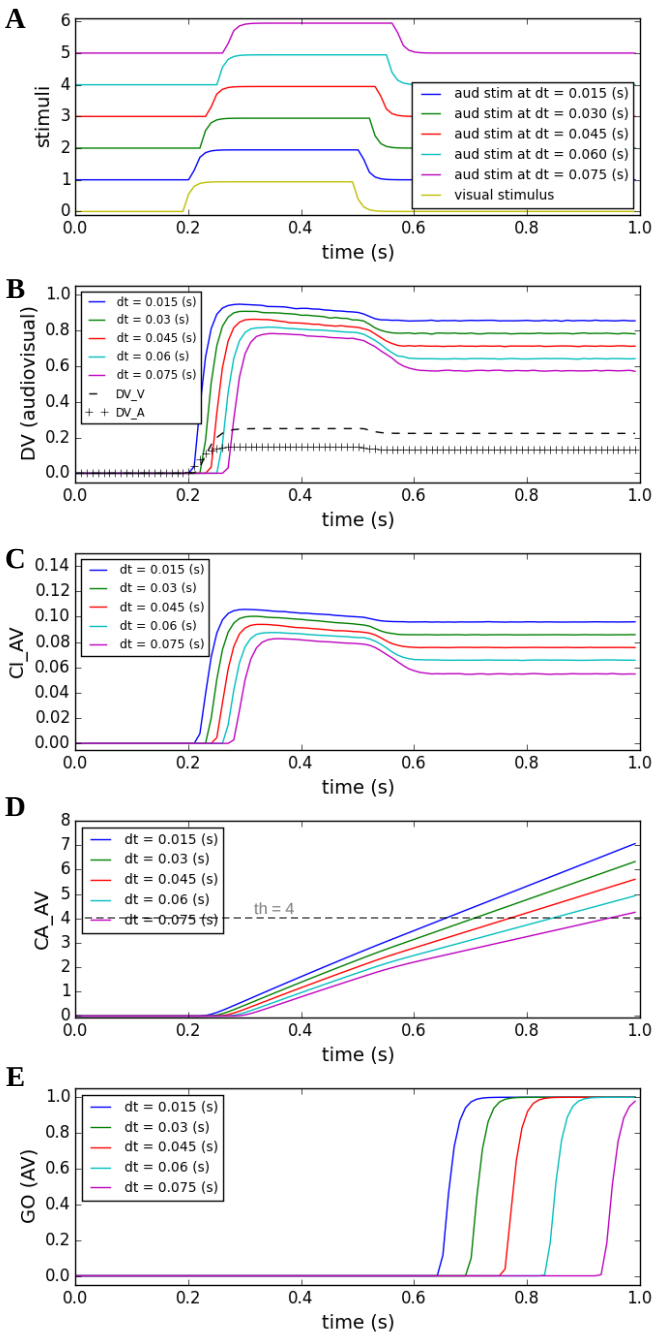

Figure 4: **Effect of temporal disparity of cross-modal stimuli on reaction time.** Five different conditions have been considered (illustrated by color coding). The visual target has a fixed position, onset time, duration, and reliability for all conditions. The position, duration and reliability of the auditory target is also fixed. **A)** The onset time of the auditory target varies changes relative to the onset time of the visual target from temporal disparity (dt) of 0.015 to 0.075 (s). **B)** The decision variable is shown being developed through time. The unimodal components are the same for all conditions. The multisensory component is changing between conditions because the spatial distance varies. **C)** The instantaneous confidence on decision is shown through time. The unimodal components are always zero because the multisensory plan is always winning because its saliency, i.e. the similarity measure, is greater than the threshold. **D)** The accumulative confidence on decision is shown as a function of time. Only the multisensory component is not zero. **E)** The GO signal is shown for execution of the winning plan. It changes from zero to one whenever the accumulative confidence passes a threshold.

3.5.4 Inverse Effectiveness

In section 5.1 we showed that for a unimodal stimulus, the reaction time decreases by increasing the reliability of the stimulus, as seen in experiments (Bell et al., 2006). However, it has been experimentally seen (Diederich and Colonius, 2004) that, for the multisensory case, this effect is reversed and the reaction time increases by increasing the reliability (reduced to intensity in our model) of the cross-modal stimuli. Here we test our model in such a task where visual and auditory stimuli are presented at fixed positions and with invariant temporal features all the time (figure 5). However, the reliabilities of the two stimuli are the same but changing within four conditions, illustrated in figure 5-A by the varying levels of blurriness of the stimuli. Figure 5-B shows the unisensory, stimulus reliabilities in the unisensory component of the decision variable (color-coded for different conditions). In all conditions the two targets are presented close enough together, in space and time, that they are perceived to be coming from the same origin, as is clear by the saliency of the multisensory plan (similarity measure in DV_AV) in figure 5-B.

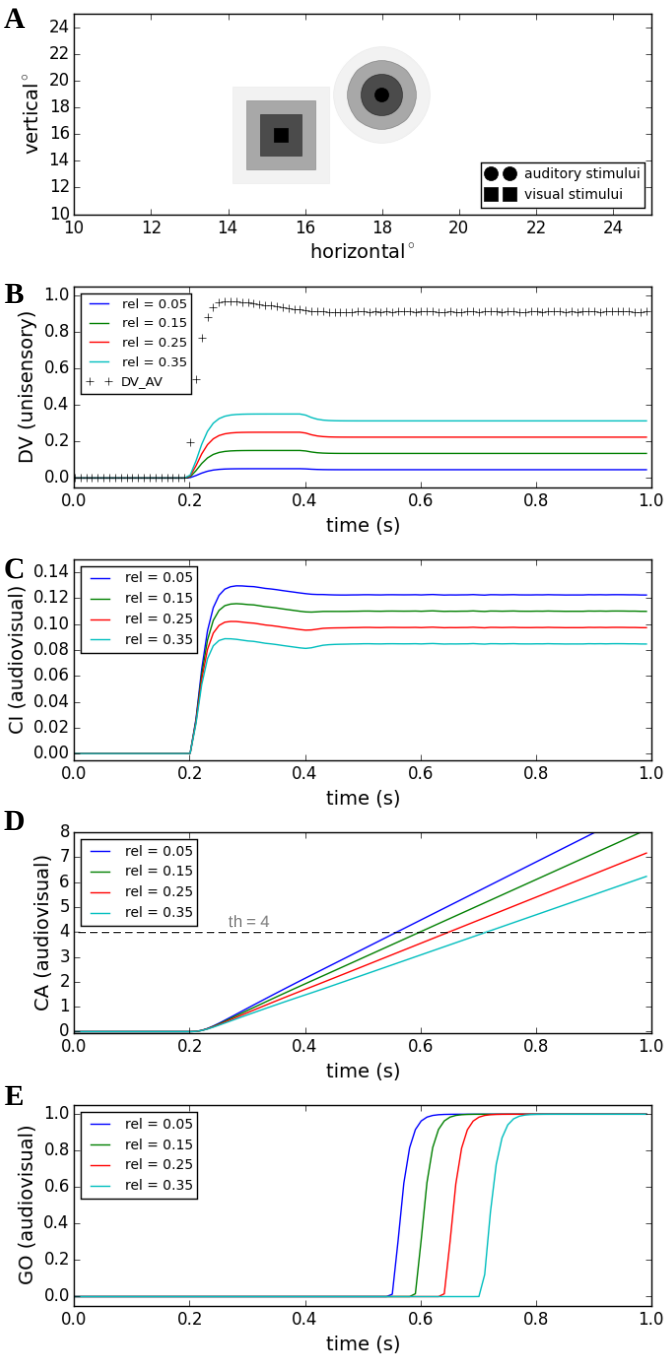

Figure 5: **Effect of the reliability of cross-modal stimuli on reaction time.** Four different conditions have been considered for all of which both visual and auditory targets have fixed positions, onset times, and durations. **A)** The reliability of the stimuli are the same for the two modalities but changing between different conditions, as illustrated by the varying levels of blurriness of stimuli. **B)** The decision variable is shown being developed through time. The multisensory component is the same for all conditions. The unimodal components are changing between conditions (while they are the same for the two modalities in one condition) because the reliabilities vary. **C)** The instantaneous confidence on decision is shown through time. The unimodal components are always zero because the multisensory plan is always winning, because its saliency, i.e. the similarity measure, is greater than the threshold. **D)** The accumulative confidence on decision is shown as a function of time. Only the multisensory component is not zero. **E)** The GO signal is shown for execution of the winning plan. It changes from zero to one whenever the accumulative confidence passes a threshold.

The reaction time in such a situation depends on how dominant the winning, multisensory plan is in the decision variable. This dominance increases when similarity measure is much greater than the unisensory target reliabilities and that happens when the intensity of the unisensory stimuli decreases (figure 5-B). This is reflected in higher instantaneous confidence on the decision for lower intensities (figure 5-C). Consequently the accumulative confidence reaches its threshold faster relatively (figure 5-D). As a result, the GO command is issued earlier for lower intensities (figure 5-E).

Thus, our model reproduces the experimental finding that reaction time is faster when the cross-modal stimuli are weaker and less intense, and consequently less reliable seen (Diederich and Colonius, 2004). The model accomplishes this by implementing the idea that the confidence on the sameness of the source of the cross-modal signals decreases if the intensity of the unimodal components increases.

3.5.5 Summary of Results

We tested our proposed mechanisms in different tasks where the spatial, temporal and reliability features of cross-modal stimuli were systematically changed. Figure 6 summarizes the preceding results by plotting reaction time as a function of the various task parameters described above. We could replicate these experimentally observed phenomena: 1) the higher the spatial disparity between the stimuli the higher the reaction time as illustrated in figure 6-A (Frens et al., 1995). 2) The higher the temporal disparity between the stimuli the higher the reaction time as depicted in figure 6-B (Frens et al., 1995). 3) If only one of the stimuli is presented, the higher its intensity, the lower the reaction time will be as explained in figure 6-C (Bell et al., 2006). 4) If two stimuli are presented close to each other in space and time, the higher their intensities the higher the reaction time will be, as illustrated in figure 6-D (Diederich and Colonius, 2004). Our model explains this wide variety of reaction times by introducing various causal structures (represented in alternative plans) for possible environmental phenomena, and suggesting that the relative confidence on a specific causal structure drives the reaction time (see Results for details, and see Discussion for more interpretations).

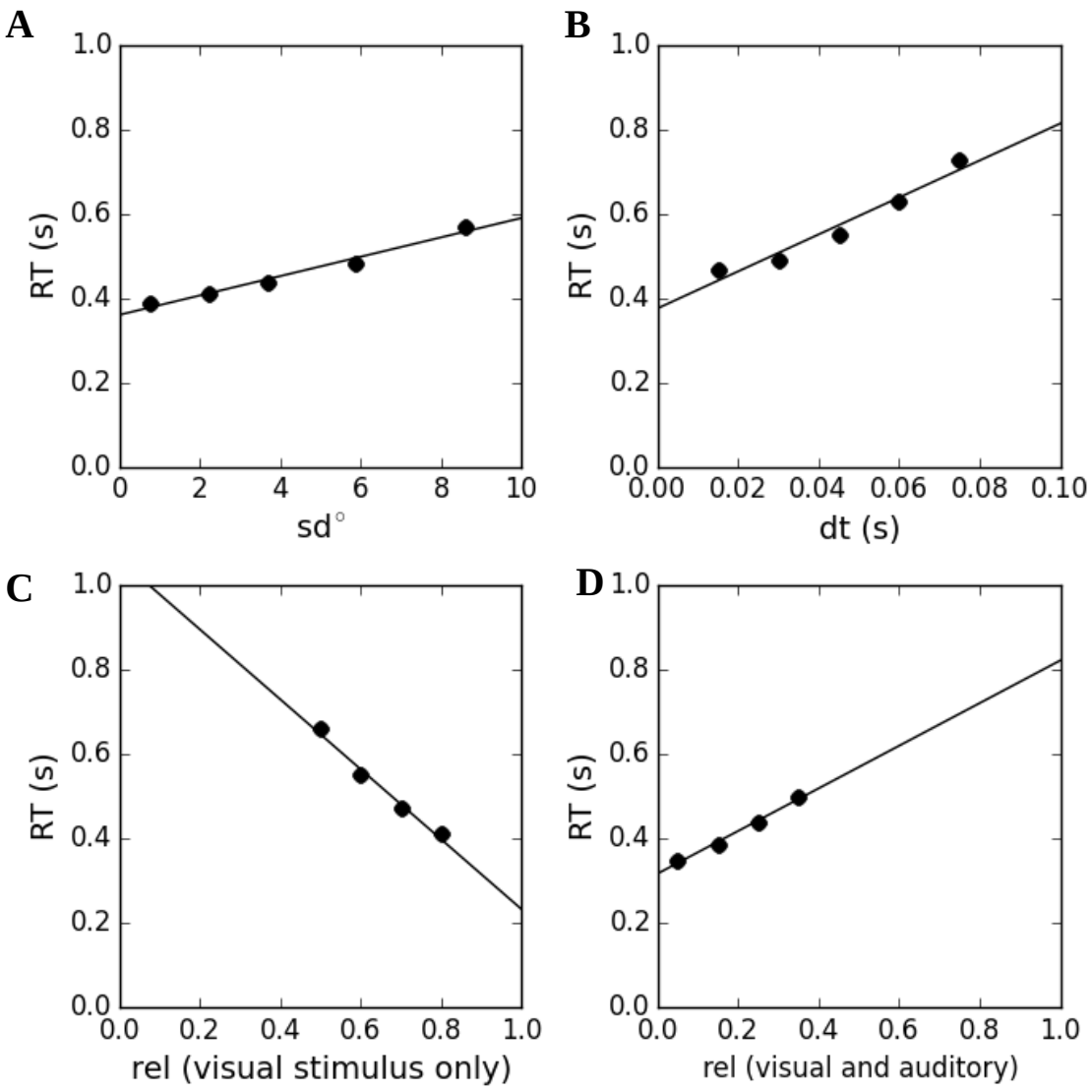

Figure 6: **Summary of the results.** Here, we want to summarize the predictions of our model for how the reaction time varies as different features of the stimuli change. **A)** The reaction time is drawn as a function of spatial disparity between the cross-modal stimuli. **B)** The reaction time is drawn as a function of the temporal disparity between the cross-modal stimuli. **C)** The reaction time is drawn as a function of the reliability of a single unimodal stimulus. **D)** The reaction time is drawn as a function of the reliabilities of the cross-modal stimuli.

3.6 Discussion

Here we proposed that the patterns of variability of saccadic reaction times (RT) towards bimodal stimuli are due to high-level cognitive processing. More specifically, the decision-making process for inference of a causal structure, and the confidence on that decision, is proposed to constitute such cognitive processing. We also consider a wider range of stimulus features including spatial, temporal and reliability aspects of the stimuli, which have shown to be affecting the reaction time (Frens et al., 1995, Bell et al., 2006). As summarized in Figure 6, this allowed us to simulate and explain that: 1) RT increases by increasing the spatial distance of the stimuli, 2) RT increases by increasing the temporal distance of the stimuli, 3) RT decreases by increasing the unimodal stimulus reliability, 4) RT increases by increasing the multimodal stimulus reliability. These findings are considered in more detail below.

3.6.1 Implications for theories of multisensory action initiation

Previous attempts to model the variability of reaction time towards bimodal stimuli assume the temporal relationships, between the presentations of the two stimuli, as the factor governing the reaction time. Either being race models that consider two separate parallel unimodal channels (Raab, 1962, Gielen et al., 1983), or the coactivation models that consider one additive stage of processing for multimodal stimuli (Schwarz, 1989, Diederich, 1992), or the time-window-of-integration models that combine the two previous ideas, they all focus on temporal processing, ignoring the spatial effect (Frens et al., 1995). They also isolate this problem from the internal cognitive processing underlying causal inference.

Our model not only considers the effects of both spatial and temporal configuration in a dynamic network, but also relates the perceptual problem of causal inference and the executive problem of action planning in a unifying framework. The model follows the more general idea that when reactionary motor responses towards sensory stimuli are avoided, we allow ourselves to plan actions based on a more complete set of information inferred from the sensory evidence (Eliasmith, 2013). Such inference extends our perception beyond the sensory information, and provides us with a wider range of action plans than sensory-driven reflexive movements (Fuster, 2005). Selection of one of such action plans

and the timing of its execution, then, depends on high-level cognitive processes, rather than reactionary sensorimotor paradigms.

The proposed cognitive system has been characterized by transformations of internal state variables through time. Time evolution of the state space is completely defined, constraining the cognitive system by the limits of planning behaviors in an uncertain, changing environment (Healy and Rowe, 2014). This dynamic nature of the signals in the model provides us with the possibility of designing new psychophysical experiments. For example, one can change the patterns of presentation of the stimuli on the time axis systematically, and see how the accumulation of evidence, through time, for and against different causal structures, change the reaction time. Or one can change the spatial and reliability features during time and see their effects on action initiation. Finally, our introduced measure of accumulative confidence can be applied to explain the reaction time other dynamic decision making tasks, where the go command is not forced by the task.

3.6.2 Significance for interpreting previous behavioral findings

It has been observed that the reaction time of planning a gaze-shift towards cross-modal stimuli is affected by the amount of spatial distance between the visual and auditory targets, and by the temporal distance between their presentations (Frens et al., 1995, Bell et al., 2005). These studies rule out statistical facilitation (Raab, 1962) by emphasizing the effect of spatial factor, besides the temporal features, on the facilitation of the reaction time. They hypothesize that the variability of reaction time is due to a multimodal stage of information processing, at a higher level than the primary unisensory processing, although they do not propose a theory for what they hypothesize.

The current model accounts for these effects, as illustrated in figures 3 and 4, by computationally and systematically realizing the intuitive idea that the confidence on the sameness of the origin of the stimuli decreases when the spatial or temporal distance between the stimuli increases. This is accomplished in two steps: 1) introduction of the spatiotemporal similarity of the multimodal stimuli as the criterion for the causal inference, and the saliency of the multisensory plan, 2) defining confidence, driving the reaction time, as how much higher the saliency of the selected plan is than the other alternatives. This meant that the plan to integrate the visual and auditory information and a gaze-shift towards

their weighted average becomes less dominant relative to unimodal gaze-shift plans, when the spatial or temporal distance between the stimuli increases. And this leads to a higher reaction time. Thus, our model proposes a cognitive theory for what previous studies hypothesized as a higher-level multisensory stage of processing.

This more general framework enables us to test the system in a wider range of tasks as well. As a first example, in gaze-shifts towards unimodal stimuli, it has been shown that the reaction time decreases by increasing the reliability (intensity) of the stimulus (Bell et al., 2006). They interpreted this as caused by reduced processing time for higher-intensity stimuli, but do not explain why the processing time decreases. Our model explains this phenomenon, as illustrated in figure 2, by the increased confidence on a unisensory gaze-shift plan, when the stimulus intensity increases. This happens because when, for example, only a visual target is present, the saliencies of the auditory and multisensory plans are zero. So, when the reliability of the visual target increases, the dominance of its corresponding gaze-shift plan increases, and the reaction time decreases.

As another example, in gaze-shifts towards multimodal stimuli presented close to each other in time and space, a reduction in reaction time has been observed when the reliabilities (intensities) of the stimuli change decrease (Diederich and Colonius, 2004). They associate this to the principle of inverse effectiveness in superior colliculus, but do not theorize a mechanism for neither of them. Our model, also, predicts that reaction time increases when the reliabilities (intensities) of the multimodal stimuli increase. This is accounted for by the relative nature of the confidence measure. With a fixed spatiotemporal configuration, and consequently a constant spatiotemporal similarity, the dominance of the multisensory plan relative to the unisensory plans decreases when the saliencies of the unisensory plans, i.e. their stimulus reliabilities, increase. And this leads to a higher reaction time.

3.6.3 Implications for neurophysiology of multisensory processing

Both the causal inference and reaction time parts of the model were designed based on the known neurophysiology about multisensory integration, working memory, decision making, gaze-shift planning and action selection. The functionalities suggested for the expert units, and the transformations realized in their connections, have been defined based

on architectural connectivity between different brain areas with known neural behavior. Sustained memory activity, contingent on action, has been shown in posterior parietal cortex (Fuster and Alexander, 1971, Cohen et al., 1997), in accord with the working memory structures in the causal inference model. The idea of multiple plan representations was inspired by laminar organization of frontal cortex (Jones et al., 1977, Canteras et al., 1990, Berendse et al., 1992, Yeterian and Pandya, 1994, Levesque et al., 1996). The inhibitive effect of decision result on plan representations was considered based on tonic inhibition of cortical and subcortical areas by the basal ganglia (Hikosaka and Wurtz, 1983b, a, Horak and Anderson, 1984).

The form in which this model is presented is a network of parallel processing units, whose states temporally change, similar to the structure of the brain. So this model of gaze-shift planning towards cross-modal stimuli can potentially be used to simulate spiking neural networks (Eliasmith et al., 2012) and then be compared to neurophysiological findings. More specifically, such a model might shed light on the mechanisms underlying the multisensory behavior of the neurons in the superior colliculus (Stein and Stanford, 2008). This might explain the spatiotemporal principles, inverse effectiveness, and unisensory behavior of SC neurons within a framework that also explains causal inference and decision making in a multisensory task.

3.6.4 Conclusion

In this paper we built up a model of action initiation, on top of our previous causal inference model. The model explained various effects on the reaction time of gaze-shifts towards cross-modal stimuli. The spatial, temporal, and reliability features of the cross-modal stimuli were systematically changed and their effects on the reaction time were reported. In accord with experimental evidence, the reaction time increased when the spatial or temporal distance between the stimuli, or their reliabilities increased.

This model introduced cognitive mechanisms, within the decision making framework of the previous model, that determine when the winning plan is sent to downstream sensorimotor machinery to be implemented (Sparks, 2002, Daemi and Crawford, 2015). Our model applied the idea of confidence on a winning plan, as the significance of that plan relative to other possible alternative plans, to control the initiation of action. Therefore,

we suggested that, in absence of a top-down command to execute the action at a fixed time, the winning plan is executed only when an accumulative measure of confidence, on a selected action plan, reaches a certain threshold.

The dynamic nature of the model allows us to predict the reaction time variability in other tasks where spatial position or reliability of the stimuli change across time, or where their temporal extensions take various forms. Also the parallel processing units in this computational model can be neurally implemented in a spiking neural network, which may help us explain firing behavior of specific brain areas or look for specific patterns of neural behavior in others.

4 A Kinematic Model for 3-D Head-Free Gaze-Shifts

Mehdi Daemi[1, 2, 3, 4], J. Douglas Crawford[1, 2, 3, 4, 5, 6 ‡]

1 Department of Biology and Neuroscience Graduate Diploma, York University, Toronto, ON, Canada

2 Centre for Vision Research, York University, Toronto, ON, Canada

3 Canadian Action and Perception Network

4 Department of Psychology, York University, Toronto, ON, Canada

5 School of Kinesiology and Health Sciences, York University, Toronto, ON, Canada

6 NSERC CREATE Brain in Action Program, York University, Toronto, ON, Canada

Front. Comput. Neurosci. 9:72. doi: 10.3389/fncom.2015.00072

Received: 21 Jan 2016; **Accepted:** 09 Jun 2016.

Edited by:

David Golomb, Ben-Gurion University of the Negev, Israel

Reviewed by:

Petia D. Koprinkova-Hristova, Bulgarian Academy of Sciences, Bulgaria

Thomas Eggert, LudwigMaximilians Universität, Germany

‡ **Correspondence:**
Dr. J. Douglas Crawford,
Center for Vision Research
Room 0009, Lassonde Bldg.
York University
4700 Keele Street
Toronto, Ontario, Canada, M3J 1P3
jdc@yorku.ca

4.1 Abstract

Rotations of the line of sight are mainly implemented by coordinated motion of the eyes and head. Here, we propose a model for the kinematics of three-dimensional (3-D) head-unrestrained gaze-shifts. The model was designed to account for major principles in the known behavior, such as gaze accuracy, spatiotemporal coordination of saccades with vestibulo-ocular reflex (VOR), relative eye and head contributions, the non-commutativity of rotations, and Listings and Fick constraints for the eyes and head respectively.

The internal algorithms of the model were inspired by known and hypothesized elements of gaze control physiology. Inputs included retinocentric location of the visual target and internal representations of initial 3-D eye and head orientation, whereas outputs were 3-D displacements of eye relative to the head and head relative to torso. Internal transformations decomposed the 2-D gaze command into 3-D eye and head commands with the use of three coordinated circuits: 1) a saccade generator, 2) a head rotation generator, 3) a VOR predictor.

Simulations illustrate that the model can implement 1) the correct 3-D reference frame transformations to generate accurate gaze shifts (despite variability in other parameters), 2) the experimentally verified constraints on static eye and head orientations during fixation, and 3) the experimentally observed 3-D trajectories of eye and head motion during gaze-shifts. We then use this model to simulate how the relative contributions of the eyes and head to vertical and horizontal gaze motion interact with constraints on torsion to influence the range of orientations of the eye in space, and the implications of these strategies for spatial version.

4.2 Introduction

Gaze-shifts, i.e. rapid reorientations of the line of sight, are the primary motor mechanism for re-directing foveal vision and attention in humans and other primates (Bizzi et al., 1971b, Tomlinson and Bahra, 1986a, Tomlinson, 1990, Guitton, 1992, Corneil and Munoz, 1996). Natural gaze-shifts in most mammals incorporate the complex coordination of eye-head movements (see Fig.2A) including a saccade towards the target, a more sluggish head movement and usually the vestibulo-ocular reflex (VOR) which keeps the eye on target during the latter parts of the head motion (Tomlinson and Bahra, 1986b, Guitton et al., 1990, Freedman and Sparks, 1997, Roy and Cullen, 1998). These components have been modeled with considerable success by several authors (Robinson, 1973, Jurgens et al., 1981, Galiana and Guitton, 1992), but the three-dimensional (3-D) aspects of gaze control have been modeled once (Tweed, 1997), and many more recently discovered properties not at all.

In the current study, we incorporate recent experimental findings into a new model for three-dimensional (3-D) gaze control, verify our mathematical approach with the use of simulations, and then use the model to explore some poorly understood aspects of eye-head coordination. In particular, we explore the interactions between the spatiotemporal rules of eye-head coordination, the 3-D constraints on eye/head orientation, and the resulting orientations of the eye (and thus retina) in space. These interactions are crucial both for understanding gaze motor coordination, and for understanding its visual consequences. Before addressing such interactions, we need to consider the basic kinematics of the eye-head gaze control system, progressing from one dimensional (1-D) to 3-D aspects.

4.2.1 Overview of Gaze Kinematics

In one dimension, gaze control kinematics reduces to the amplitudes and temporal sequencing of eye and head motion (Tomlinson and Bahra, 1986b, Guitton and Volle, 1987, Guitton, 1992, Sparks et al., 2002). The typical sequence of events includes a saccade, followed by a slower head movement and a compensatory vestibuloocular eye movement (Figure 3). The aspects of this progression that we will explore here include the variable timing of saccade, head movement and VOR, the influence of initial eye and head

orientations, relative magnitudes of the contribution of these different phases to the gaze-shift and where the head falls in space after the gaze-shift.

Additional complexity emerges when one considers gaze-shifts from a two-dimensional (2-D) perspective. For example, the eye and head provide different relative contributions to horizontal and vertical gaze motion, which must be predictably accounted for saccades to produce accurate gaze shifts (Freedman and Sparks, 1997, Goossens and Van Opstal, 1997), and for the eye and head to end up in the right positions after the VOR (Crawford and Guitton, 1997b, Misslisch et al., 1998).

Finally, gaze control reaches its highest degree of complexity in 3-D (Glenn and Vilis, 1992, Crawford et al., 2003). First, there is an added dimension of motion control: torsion, which roughly corresponds to rotations of the eyes and/or head about an axis parallel to the line of sight pointing directly forward. Torsion influences direction perception for non-foveal targets (Klier and Crawford, 1998), binoclular correspondence for stereo vision (Misslisch et al., 2001, Schreiber et al., 2001), and must be stabilized for useful vision (Crawford and Vilis, 1991, Fetter et al., 1992, Angelaki and Dickman, 2003). More fundamentally, a 3-D description requires one to account for the non-commutative (order-dependent) properties of rotations (Tweed and Vilis, 1987, Hepp, 1994). These properties influence not only ocular torsion and the degrees of freedom problem, but also gaze accuracy, for reasons related to reference frame transformations.

The location of a visual stimulus is initially described in an eye-centered reference frame by the pattern of light that falls on the retina and the resulting activation of eye-fixed photoreceptors (Westheimer, 1959). Whereas the orientation of the eye and the motor commands for its movement are encoded in a head-centered reference frame (Crawford and Vilis, 1992a, Crawford, 1994) while head orientation and head movements are encoded in a coordinate system attached to the torso (Klier et al., 2007). This is because the eye muscles which move the eyes are fixed to the head and the neck muscles which move the head are fixed to the shoulder, although the dynamic actions of the muscles are also modulated by the orientations of the bodies that they control (Farshadmanesh et al., 2007). Given the descriptions of these signals in different reference frames, sensory-driven planning of gaze-shifts consists in a structured set of reference frame transformations

(Sparks and Mays, 1990b, Klier et al., 2001). This is often circumvented in 1-D and 2-D models of gaze-shift that borrow the math of the translational motion to approximate rotation, but when the full properties of 3-D rotation are realistically incorporated into such models, reference frame transformations cannot be avoided. Instead, they must be embedded in the fundamental structure of the model (Crawford and Guitton, 1997d, Tweed, 1997, Blohm and Crawford, 2007). This too will be incorporated into our model and simulated (Figure 4).

Another factor to consider is that biological constraints that limit the degrees of freedom of the range of eye and head orientations to a subset of their mechanically possible range (simulated below in figures 5 and 6). Suppose an arbitrary rotating rigid body (whose orientation is defined as the rotation which moves it from a set reference direction to face another specific direction) is described in a fixed coordinate system. Different patterns, or even combinations, of rotations can possibly bring the rigid body onto a specific direction. If the rigid body obeys Donders' law, there is an injective map between the domain of the directions (3-D vectors) and the domain of the orientations (3-by-3 rotation matrices), i.e. each time the rigid body faces in a particular direction, it only assumes one 3-D orientation (Glenn and Vilis, 1992, Crawford et al., 2003). Orientation of the eye relative to the head and orientation of the head relative to the shoulder obey Donders' law between gaze-shifts when the head and body are normal upright postures (Misslisch et al., 1994, Klier and Crawford, 2003). Orientation of eye-in-head has also been shown to obey the Listings' law (Ferman et al., 1987b, a, Tweed and Vilis, 1990, Straumann et al., 1991); If torsion is defined as rotation about the axis parallel to gaze at the primary eye position, then Lesting's law states that eye orientation always falls within a 2-D horizontal-vertical range with zero torsion known as Listing's plane. Orientation of head-on-shoulder has been shown to obey the Fick strategy (Glenn and Vilis, 1992, Crawford et al., 1999b, Klier et al., 2007); where torsion is constrained to be zero in Fick coordinates described as a sequence of three successive rotations about vertical (fixed in the body), horizontal (mounted on the vertical axis) and torsional (mounted on the first two, i.e. fixed in the head) axes. Mechanical factors appear to aid these constraints by implementing some of the position-dependencies required to deal with non-commutitivity (Demer et al., 2000, Ghasia and Angelaki, 2005, Klier et al., 2006). But ultimately mechanical factors cannot enforce these constraints

without blocking torsion altogether. On the contrary, these constraints are violated (e.g. leading to large torsional rotations) whenever required by other behavioral circumstances (Misslisch et al., 1998, Crawford et al., 1999b).

Note that these systems seem to be primarily concerned with enforcing Donders' law during fixations at the end of the gaze-shift when both the eye and head are relatively stable, perhaps because of their various implications for sensory perception. Listing's law is also obeyed during saccades with the head-fixed (Ferman et al., 1987b, Tweed and Vilis, 1990). However, when the head is free to move, both the eye (Crawford and Vilis, 1991, Crawford et al., 1999b) and head (Ceylan et al., 2000) are known to depart from their Donders' ranges during gaze movement, for reasons that will be described below and simulated in figures 6 and 7. This also suggests additional aspects of neural control that, to date, have only been considered for the eye.

Thus, a complete model of the head-free gaze-shifts needs to incorporate both the reference frame transformations and some solution to the behavioral constraints described above. Further, such a model should plan for spatial and temporal coordination of saccade, head movement and VOR. Furthermore, variability of the contribution of head movement to the gaze-shift, the variability of the sizes of saccade and VOR and the variability of these contributions in different spatial directions have to be considered. These factors interact in complex fashions (Figures 5, 8, 9) that have only partially been explored. Again, this remains an important topic, because it has fundamental implications for both vision and motor control. But before attempting to address this goal, we will briefly review previous attempts to model gaze control, ranging from early models of the 1-D saccade system to the most recent 3-D model of eye-head coordination.

4.2.2 Gaze Control Models: from 1-D Saccades to 3-D Eye-Head Control

Attempts to model the gaze control system have generally advanced from 1-D models of head-restrained saccades towards multi-dimensional models of head-unrestrained gaze-shifts. The first models of gaze-shift were dynamic models of one-dimensional head-fixed saccades. Robinson (1973) assumed that saccades are driven by a fast feedback loop allowing trajectory corrections on the fly (Robinson, 1973). Jurgens et.al (1981) observed that despite the variability of the duration and speed of the saccades their accuracy is almost

constant, and considered this observation favoring the hypothesis of local feedback (Jurgens et al., 1981). The next question addressed was if the 1-D saccade models could be generalized for oblique and 3-D saccades. Van Gisbergen et.al (1985) observed for oblique saccades that the horizontal and vertical components of the movement start simultaneously and are adjusted relatively such that straight trajectories are produced (van Gisbergen et al., 1985). Then they found that a model based on a common source of motor command for horizontal and vertical components agrees with the data rather than a model based on independent 1-D motor commands for the two components. In parallel to this, many of these principles, combined with models of the VOR, were incorporated into models of eye-head gaze control. For example, Bizzi et.al (1973) developed this idea that the head movement during gaze-shift attenuates the saccade amplitude by an amount equal to the VOR (Morasso et al., 1973). Galiana et.al (1992) proposed a kinematic model of eye-head coordination in one dimension, in which they introduced the idea of VOR gain changing as a function of gaze-shift amplitude (Galiana and Guitton, 1992).

The development of 3-D models of gaze-shifts followed a similar course, but shifted forward by a decade. Tweed and Vilis (1987) mathematically proved, through non-commutativity of 3-D rotations, that the 3-D saccades should be planned based on 3-D kinematics of the eye rather than linear generalization of the 1-D models (Tweed and Vilis, 1987). Subsequent 3-D models of the saccade generator either focused on the question of eye muscle contribution to Listing's law (Quaia and Optican, 1998, Raphan, 1998), reference frame transformations for saccades (Crawford and Guitton, 1997c), or interactions between saccades and vestibular system (Crawford et al., 2011). Tweed (1997) proposed the first (and to date only) three-dimensional kinematic model of eye-head saccadic system (Tweed, 1997). This model starts by selecting a specific desired 3-D orientation of the eye-in-space and then the eye and head are driven by a dynamic gaze error signal and he defined the constraints on the velocity signals, i.e. the head, for instance, is always rotating around Fick axes. Some aspects of Tweed's framework have since been used for modeling other vision-based goal-directed actions (Blohm and Crawford, 2007), but otherwise the advance of theoretical models of 3-D eye-head gaze control seems to have halted for the past 17 years.

In contrast, our knowledge of the physiology and behavior of 3-D gaze control has grown considerably in the past 17 years. For example, a 3-D analysis of head-unrestrained gaze shifts evoked by stimulation of the superior colliculus (SC), frontal eye-fields (FEF) and supplementary eye fields (SEF) suggests that these structures encode desired 2-D gaze direction, rather than desired 3-D eye orientation in space (Martinez-Trujillo et al., 2003, Monteon et al., 2010). Based on the experimental literature, it appears that elaboration into 3-D commands does not happen until the level of the brainstem reticular formation, likely with aid from the cerebellum, and occurs separately for eye and head orientation signals (Klier et al., 2003).

4.2.3 Aims of the Current Study

In this paper, we are proposing a model for the kinematics of 3-D head unrestrained gaze-shifts towards visual targets. Our motivation for this study is 1) that a number of experimental advances in understanding 3-D eye-head gaze-shifts have occurred in the past 17 years that were not considered or incorporated into Tweed's 1997 model, 2) we wished to build a kinematic framework to inspire further experiments, modeling approaches (e.g. neural network studies), and data interpretation, and 3) we wished to use this model to explore several questions that have largely been overlooked (except in thought experiments) in the 3-D gaze control literature, in particular how the various aspects of control described above interact and specially how they influence eye orientation, and thus the orientation of visual stimuli in space relative to the retina.

In brief, we apply the behavioral constraints of eye and head on their respective final orientations such that their instantaneous orientations during the gaze-shift do not necessarily obey these rules. Instantaneous orientations of eye and head can be derived assuming that the axis of rotations remain constant during the movements. Desired head position is assumed to be dependent on the desired gaze position. We have defined two parameters that control this dependence in horizontal and vertical directions. The larger the values of these parameters the closer the head would fall to the desired gaze direction. For very small values of these parameters, the model is reduced to an experimentally-verified model for the head-fixed gaze-shifts obeying all geometrical constraints for saccade. Unlike the eye, one unique head rotation is planned for the gaze-shift. A parameter has

been defined to break this single head rotation into a part which contributes to gaze and one which is cancelled out by VOR. This parameter subsequently determines the amount of VOR eye movement as well. The model parameters are independent of the model structure and constraints. Assigning different values to these parameters, we can plan various patterns of eye-head coordination for making a gaze-shift. The mathematical formalisms of this model are described in the next section, followed by simulations designed to test the success of the model and then extend it to predict/interpret new situations.

4.3 Model Formulation

4.3.1 Overview

In mechanics, for determination of mechanical behavior of a component, the classic approach is to solve the governing laws of conservation (of mass, momentum and energy) for a specific geometry, material and initial / boundary / loading conditions. In this approach one can find general analytic solutions for the distribution of motion (displacement, velocity and acceleration), deformation and stress. However, this approach is not applicable to most real practical situations. Therefore, we have used an alternative approach that is usually used in "engineering design". This approach includes three levels: the first level, static kinematic model, includes deriving the desired positions and patterns of motion for different components in the plant for meeting a kinematic end. The second level, temporal discretization, includes determination of a time-framework and associating specific temporal growth functions to different desired motions and then deriving the velocities and accelerations of different components as functions of time. The third level includes putting the known kinematic variables, external loads and the mechanical properties of the plant into the equations of conservation of momentum and solving them for the unknown force / torque functions. In this paper, we describe our model at the first level of this general approach: a static kinematic model for 3-D head-unrestrained gaze-shifts towards visual targets.

Figure 1 shows a summary of the signals in the model and their relations with each other. The small red and blue boxes are inputs and outputs of the system respectively. Each signal is mathematically computed from its input signals. The major internal computations can be

divided into three groups: one group responsible for calculating the total head rotation (large green box), one responsible for predicting the VOR-related eye rotation (large violet box), and one responsible for calculating the saccadic-related eye rotation (large red box).

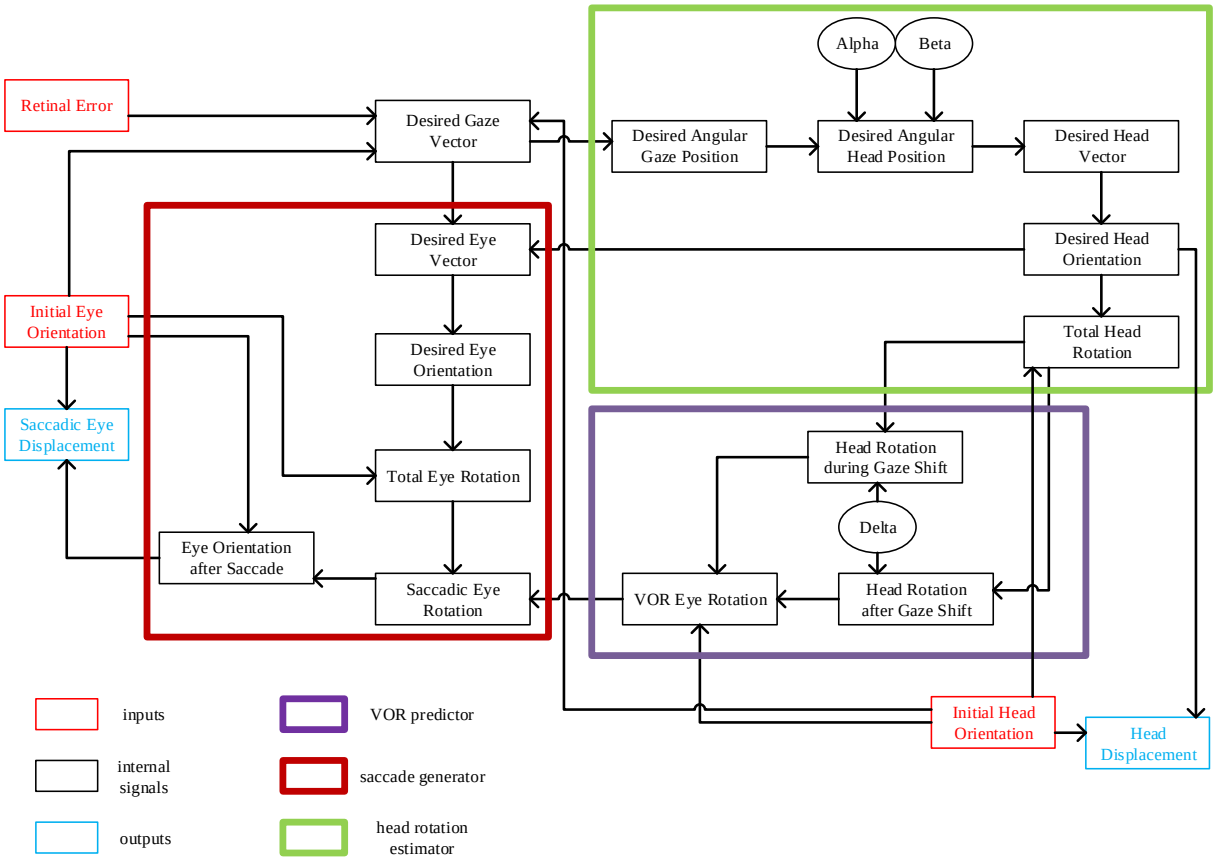

Figure 1: **Flow of Information in the Static Kinematic Model**. Red and blue rectangles show model inputs and outputs, respectively. Black ovals are the model parameters. Big thick red box shows the part of the model involved in computation of the saccadic eye movement. Big thick green box shows the part of the model which computes the head movement. Big thick violet box shows the VOR predictor. Each signal is computed from the signals that have inputs to it.

This sequence of calculations begins when light is emitted from a target in the periphery onto the retina and the sensory signal to drive the gaze-shift is constructed as retinal error, the eye-centred 2-D vector which characterizes the distance and direction of the retinal image of the target relative to the fovea. In our model, this is geometrically equivalent to gaze 2-D motor error in retinal coordinates, and thus could represent spatial activity in the brain at any point from the retina to the superior colliculus (Klier et al., 2001, DeSouza et al., 2011). Desired gaze (eye-in-space) vector, a unit vector directing towards the target, is calculated from retinal error and the internal knowledge of the initial 3-D orientations of eye-in-head and head-on-shoulder, which could be derived from proprioceptive signals (Steinbach, 1987, Wang et al., 2007) and / or efference copies from 'neural integrators' in the brainstem (Cannon and Robinson, 1987, Crawford et al., 1991, Farshadmanesh et al., 2007). Note that this gaze vector does not yet specify torsion of the eye in space; it is intermediate computational stage useful in decomposing retinal error into both eye and head components (see below). Thus, the initial stages of the model is based on experimental observations that early gaze centres specify 2-D direction, with implementation of 3-D eye and head constraints further downstream (van Opstal et al., 1991, Klier and Crawford, 2003).

In order to calculate the desired head movement (Fig. 1; green box), the desired gaze vector is first converted into angular gaze position, a 2-D version of desired gaze vector in spherical coordinates. Desired angular gaze position, a 2-D version of desired gaze vector in spherical coordinates, is then calculated. Desired angular head position is computed from the desired angular gaze position and two of the model parameters: α, β. These two parameters, α & β, have been defined to determine where the head falls relative to the gaze in horizontal and vertical directions respectively. The 3-D desired head vector is computed from the 2-D desired angular head position. Desired head orientation that conforms to the Fick strategy (zero torsion in Fick coordinates) is then calculated from the desired head vector. Knowing the initial and desired head orientations, the total head rotation is calculated, and then converted into a head displacement command (see below for physiological interpretation of this output).

In order to generate a saccade that is correctly coordinated with head movement (Crawford et al., 1999b), our model first predicts the VOR eye movement that will occur toward the end of the movement (Fig. 1; violet box). This is not as difficult as it might sound. Assuming the constancy of the axis of head rotation throughout the gaze shift, the total head rotation is broken down into two parts with the aid of one of the model parameters, δ. This parameter defines two phases of the head rotation; a first one which contributes to the gaze-shift and a second one which is cancelled out by vestibulo-ocular reflex (VOR). Then, knowing the initial head orientation and the two parts of head rotation, then one can predict the ideal VOR eye movement that would stabilize 3-D gaze orientation during the second phase of the head rotation. This is not the same physiological mechanism as the actual VOR (which is driven by signals from the semicircular canal), but in our simulations we assume an ideal VOR model and use the same signal. In real world conditions this behavior would occur thousands of times each day, and thus provide ample opportunity to train a dynamic neural network to learn the calculations described here. The physiological basis for this hypothetical network could involve the brainstem and cerebellum.

The last part of the model is involved in computing the 3-D saccade vector (Fig. 1; Red box), meaning a saccade that also includes the torsional components required to offset the oncoming VOR (Crawford et al., 1999a). Having computed the desired head orientation and desired gaze vector, we first calculate the desired final 2-D eye direction vector relative to head (after saccade and VOR). We then covert this into desired eye orientation (after the saccade and VOR) to fall in the Listing's plane. Knowing the initial and desired eye orientations, we calculate the total eye rotation. Having computed the total eye rotation and the VOR eye rotation, we can finally calculate the saccadic eye rotation. This rotation not only helps foveate the target but also compensates for all VOR components in a predictive fashion. This is then converted into the desired final eye orientation after the saccade, and initial eye orientation is subtracted from this to produce desired 3-D eye displacement in Listing's plane coordinates, the command that is mathematically appropriate to drive the known 3-D coordinates of premotor oculomotor structures (Crawford and Vilis, 1992a, Crawford, 1994), and henceforth derivatives of eye orientation coded within the phasic burst of motoneurons (Ghasia and Angelaki, 2005, Klier et al., 2006, Farshadmanesh et al., 2012a). The torsional component of this displacement command might be generated by the

nucleus tegmenti reticularis pontis (Van Opstal et al., 1996), eventually leading to activation of the torsional burst neurons. Thus, these parts of the model reflect what might happen in the real brain between the superior colliculus (Klier et al., 2001)and the oculomotor burst neurons (Henn et al., 1991, Crawford and Vilis, 1992b, Crawford, 1994).

Very little is known about the mathematical details of brainstem and spinal motor commands for the head, but they appear to follow similar principles to that seen in the oculomotor system (Klier et al., 2007, Farshadmanesh et al., 2012a). Therefore, to model the final output of our head control system we also subtracted initial 3-D eye orientation from desired 3-D eye orientation to obtain a 3-D dispacement command. Note that for such displacement outputs, it is necessary that any further position-dependences, such as the half-angle rule of eye velocities for Listing's law, are implemented further downstream, likely at the level of muscles (Demer et al., 2000, Ghasia and Angelaki, 2005, Klier et al., 2006, Farshadmanesh et al., 2012a, Farshadmanesh et al., 2012b).

4.3.2 Basic Mathematical Framework

As illustrated in figure 2, *eye vector* (red) is a vector fixed to the eye ball aligned from the center of the eye ball to the fovea. Assuming the head as a sphere, *head vector* (green) is a vector fixed to the head, aligned from the center of this sphere to the nose. Initially, eye vector intersects with the screen at the initial fixation point. Gaze-shift is to be planned to foveate the desired target, i.e. move the eye vector to intersect the desired target location on the screen. This shift of eye vector is executed by a coordinated pattern of eye and head movements.

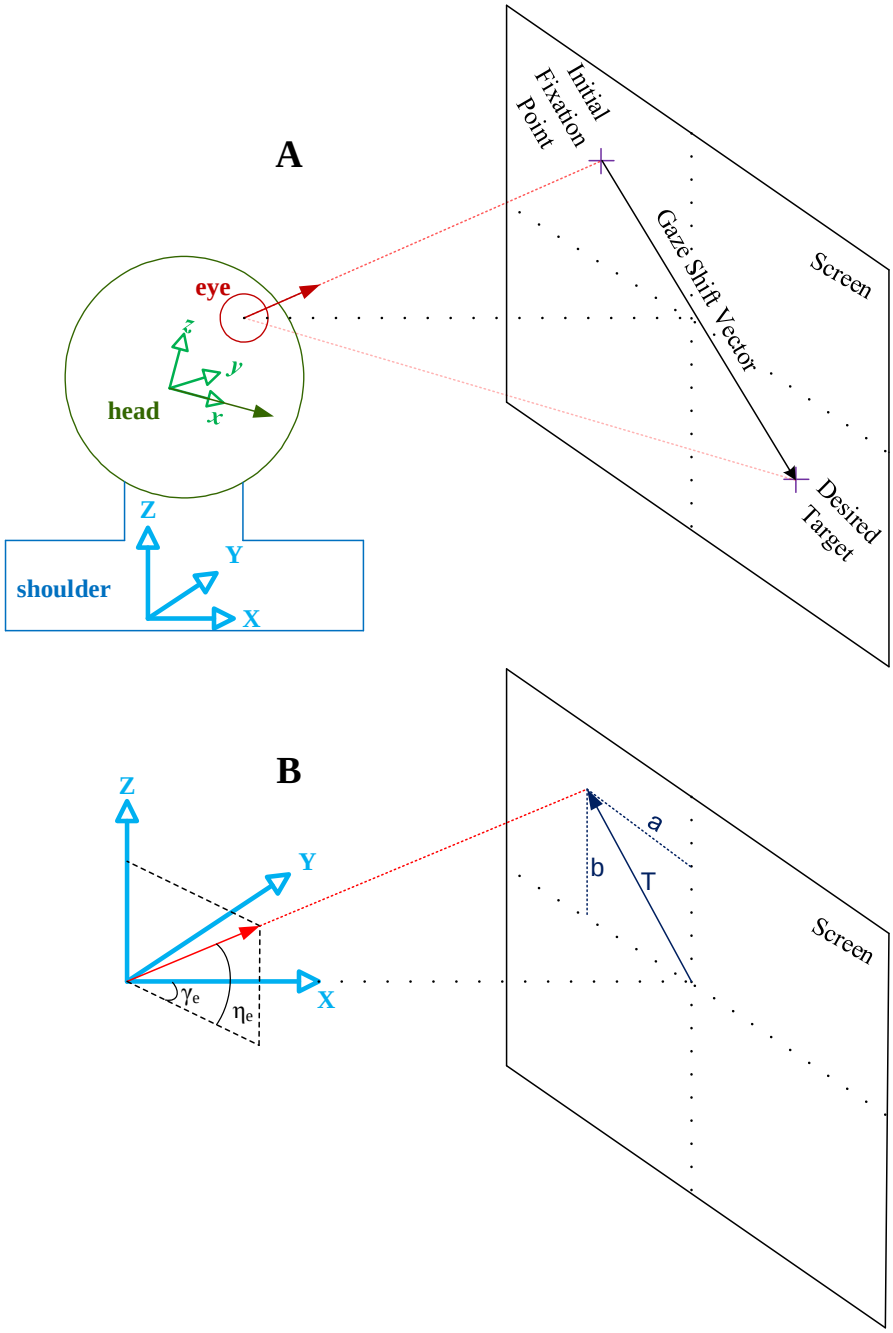

Figure 2: **Illustration of the Geometrical Framework for Studying Head-Free Gaze-Shift. (A)** Head coordinate system, shown by the green axes fixed to the head, explains everything relative to the head. Shoulder or space coordinate system, shown by the blue axes fixed to the shoulder, explains everything relative to the space. Green vector is the head vector which is fixed to the head and moving with it. Red vector is the eye vector which connects center of the eye ball to the fovea. In the reference condition eye and head vectors are aligned in the same direction and intersect with the center of the screen. Eye vector, defined in head coordinate system, is called eye-in-head vector and eye vector, defined in space coordinate system, is called gaze vector. Head vector is defined only relative to space. **(B)** Space coordinate system is drawn again to show how eye vector is characterized in space to represent the gaze vector. Gaze vector, or eye vector in space coordinates, is a unit vector which shows where the eye is fixating. Gaze vector can have a 2-D angular representation based on the angles it creates in spherical coordinates with the axes (the same applies to the head vector with angles not shown here). Gaze vector can be derived if we know where on the screen the subject is fixating, which is characterized by vector.

As illustrated in figure 2, we define a coordinate system attached to the shoulder and fixed to the space. $\{X, Y, Z\}$ of this so-called space coordinate system are respectively orthogonal to the coronal, sagittal and axial anatomical body planes. We also define a coordinate system attached to the head which moves with the movement of the head. We define reference condition as the straight-ahead configuration of eye and head where $\{x, y, z\}$ of the head coordinate system is aligned with $\{X, Y, Z\}$ and eye vector is aligned with x and X. For instance, in a conventional experimental setup for eye movement research, where the subject is sitting in front of a screen, reference condition is typically when the subject is fixating the center of the screen and eye vector and head vector are parallel.

Eye vector is called eye-in-head vector, $\vec{e}$, when defined in head coordinate system and is called gaze vector, $\vec{g}$, when defined in space coordinate system. Head vector, $\vec{h}$, is only defined relative to space coordinate system. For any configuration of oculomotor system, eye-in-head orientation, $\boldsymbol{E}$, head orientation, $\boldsymbol{H}$, and gaze orientation, $\boldsymbol{G}$, are rotation matrices which rotate $\vec{e}$, $\vec{h}$, $\vec{g}$ respectively, from the reference condition to their current configuration (letters "r", "i" and "d" as subscripts, denote reference, initial and desired conditions):

$$\vec{e} = \boldsymbol{E} \times \vec{e}_r \tag{1}$$

$$\vec{h} = \boldsymbol{H} \times \vec{h}_r \tag{2}$$

$$\vec{g} = \boldsymbol{G} \times \vec{g}_r \tag{3}$$

At any arbitrary configuration, if we rotate the eye-in-head vector by head orientation matrix we will derive the gaze vector. So, gaze orientation is always the multiplication of head and eye-in-head orientations:

$$\boldsymbol{G} = \boldsymbol{H} \times \boldsymbol{E} \tag{4}$$

We define $\vec{c}$, the 2-D angular gaze position and $\vec{b}$, the 2-D angular head position based on the defining angles of the eye and head vectors in the spherical version of the space coordinate system (these angles are shown for eye vector in figure 1-B. the same applies for the head vector.)

$$\vec{c} = \left[\frac{\pi}{2} - \gamma_e \; ; \; \frac{\pi}{2} - \eta_e\right] \tag{5}$$

$$\vec{b} = \left[\frac{\pi}{2} - \gamma_h \; ; \; \frac{\pi}{2} - \eta_h\right] \tag{6}$$

Gaze and head vectors can be directly derived from the spherical angles:

$$\vec{g} = [\sin\eta_e \,.\, \sin\gamma_e \;\; ; \;\; \sin\eta_e \,.\, \cos\gamma_e \;\; ; \;\; \cos\eta_e] \tag{7}$$

$$\vec{h} = [\sin\eta_h \,.\, \sin\gamma_h \;\; ; \;\; \sin\eta_h \,.\, \cos\gamma_h \;\; ; \;\; \cos\eta_h] \tag{8}$$

We also define the target position on the screen by the vector $\vec{T} = [a; b]$ as it is illustrated in figure 1-B. If "t" is the distance between the eye and the center of the screen, $\vec{T}$ and $\vec{g}$ can be derived from each other:

$$\vec{g} = \frac{1}{\sqrt{t^2 + a^2 + b^2}}[t \; ; \; a \; ; \; b] \tag{9}$$

$$\vec{T} = t \times [g(2) \; ; \; g(3)] \tag{10}$$

The main input of the oculomotor system is supposed to be the retinal error. In our formulation, we define a 3-D version of this signal, $\vec{g}_{RE}$, as the desired gaze vector relative to the initial gaze orientation:

$$\vec{g}_{RE} = \boldsymbol{G_i}^{-1} \times \overrightarrow{g_d} \tag{11}$$

A 2-D angular version of this signal can also be derived from the previous vector:

$$RE = [\cos^{-1}(g_{RE}(3)) \; ; \; \cos^{-1}(\frac{g_{RE}(2)}{\sin(\cos^{-1}(g_{RE}(3)))})] \tag{12}$$

4.3.3 Motor Mechanisms of Eye-Head Movement

There are three distinct motor mechanisms that move the effectors (eye and head). For planning a gaze-shift, the brain has the luxury of choosing an arbitrary combination of these three mechanisms by determining the amount of their contribution and the pattern of their temporal implementation. The subject is initially fixating an arbitrary target and orientations of eye and head at initial condition are known variables of our problem:

$$\vec{e}_i = \boldsymbol{E_i} \times \vec{e}_r \tag{13}$$

$$\vec{h}_i = \boldsymbol{H_i} \times \vec{h}_r \tag{14}$$

$$\vec{g}_i = \boldsymbol{H_i} \times \boldsymbol{E_i} \times \vec{g}_r \tag{15}$$

Saccade

Saccade is the movement of eye relative to the head. Eye rotates in the head by rotation matrix $\boldsymbol{Re}$ and head stays fixed:

$$\vec{e} = \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{e}_r \tag{16}$$

$$\vec{h} = \boldsymbol{H_i} \times \vec{h}_r \tag{17}$$

$$\vec{g} = \boldsymbol{H_i} \times \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{g}_r \tag{18}$$

Eye-Carrying Head Rotation

Head is driven towards the target while no motor command is sent to eye muscles. Head rotates, moving eye with itself such that eye-in-head position remains unchanged (Guitton et al., 1984). Head and eye rotate together by unknown rotation matrix $\boldsymbol{Rh}$:

$$\vec{e} = \boldsymbol{E_i} \times \overrightarrow{eh}_r \tag{19}$$

$$\vec{h} = \boldsymbol{Rh} \times \boldsymbol{H_i} \times \vec{h}_r \tag{20}$$

$$\vec{g} = \boldsymbol{Rh} \times \boldsymbol{H_i} \times \boldsymbol{E_i} \times \vec{g}_r \tag{21}$$

Gaze-Stabilized Head Rotation

This is the arbitrary movement of the head while the gaze is fixated. Head rotates while eye rotates in head to compensate for head movement and keep the gaze stabilized. This type of eye movement is called vestibulo-ocular reflex. While head is rotating by unknown rotation matrix $\boldsymbol{Rw}$, eye is moving in the opposite direction by rotation matrix $\boldsymbol{Rv}$:

$$\vec{e}_d = \boldsymbol{Rv} \times \boldsymbol{E_i} \times \overrightarrow{eh}_r \tag{22}$$

$$\vec{h}_d = \boldsymbol{Rw} \times \boldsymbol{H_i} \times \vec{h}_r \quad (23)$$

$$\vec{g}_d = \boldsymbol{Rw} \times \boldsymbol{H_i} \times \boldsymbol{Rv} \times \boldsymbol{E_i} \times \vec{g}_r \quad (24)$$

4.3.4 Static Kinematic Model

As it is experimentally observed and schematically illustrated in figure 3-A, the gaze-shift typically begins when the saccadic eye movement rapidly changes the positions of the eyes relative to the head and it ends when the line of sight is directed toward the visual target, and the rapid eye movement component of the gaze-shift ends at approximately the same time. The head continues moving towards the target while the eyes move in the opposite direction at a velocity that is approximately the same as that of the head. As a result, the direction of the line of sight changes very little (Bizzi et al., 1971b, Bizzi et al., 1972, Zangemeister et al., 1981).

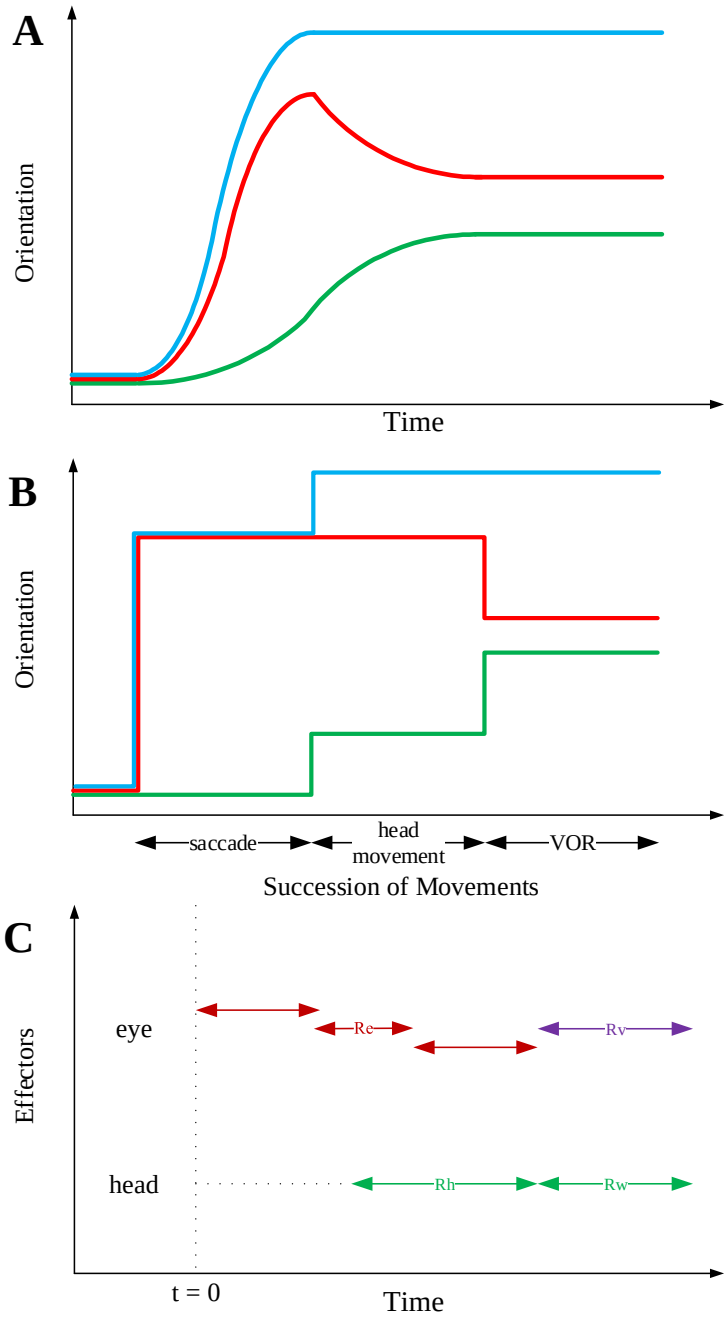

Figure 3: **Sequential Structure of Rotations in the Kinematic Model**. In the first two panels, blue, red and green curves respectively depict gaze, eye-in-head and head trajectories. **(A)** Typical 1-D behavioral diagram from the experiments on natural head-unrestrained gaze-shift (Guitton et al., 1990, Freedman and Sparks, 1997). This observed pattern has inspired the sequence of events devised in the static kinematic model. **(B)** Succession of movements in the kinematic model. Head remains fixed while the eye is moving in the head. Then, head rotates, moving eye with itself such that eye-in-head position remains unchanged; this rotation foveates the target. Then, head rotates to its definite position, while eye rotates in head to compensate for head movement and keep the target foveated. **(C)** Having solved the equations of the model based on our physiologically inspired assumptions and constraints, we find that the saccadic eye movement has its independent axis and can be implemented in any duration of time which ends before onset of VOR (red double-headed arrows). Onset of head movement is arbitrary but its two parts are implemented continuously after each other (green double-headed arrows). Eye rotation during VOR is implemented right at the time when the second part of head movement is happening (violet double-headed arrow).

According to observations of visual orienting behavior it is clear that movements of the eyes and head can begin at approximately the same times. However, recording the activity of neck and eye muscles reveals that even when movement onsets are synchronous, the command to move the head precedes the command to move the eyes (Bizzi et al., 1971a, Zangemeister et al., 1981). Furthermore, inspection of the behavior over a broad range of gaze-shift amplitudes, task requirements, and target predictability indicates that the relative timing of eye and head movements is variable (Zangemeister and Stark, 1982, Guitton and Volle, 1987, Freedman and Sparks, 1997, Crawford et al., 1999b): the head can lag the onset of eye movements during small amplitude gaze-shifts, but during large amplitude movements, or movements to target locations that are predictable head movements can begin well before saccades. Electrical stimulation in the omnipause neuron region can delay saccade onset without altering the initiation of head movements (Gandhi and Sparks, 2007); evidence that the triggering mechanisms for the eyes and head are not shared.

According to the evidence about temporal coupling of eye and head movements described above, a separation (at least with respect to movement initiation) of head and eye command signals can be identified within the brainstem structures that control coordinated eye–head movements. This may indicate that the brain plans a gaze-shift at different levels. Accordingly, inspired by the fundamentals of engineering design, we propose that a complete model of gaze-shift is planned in three levels of information processing. At a higher level, illustrated in figure 3-B, we propose a succession of movements as a structure for computing the motor commands for eye and head. At a middle level, sketched in figure 3-C, a temporal structure for implementation of these movement commands should be proposed. It can be shown that these two levels are independent, i.e. succession used for computation of motor commands does not dictate the timing of their implementation. Rather, saccadic eye movement can start before or after the onset of head movement and can finish well before the onset of VOR. At a lower level, the required torques are calculated by putting the then-known kinematic variables in governing conservation equations, and then, knowing the structure of motoneurons and muscles, the required neural signals could be derived. Having emphasized this hierarchical structure, in this paper, we are only concerned with the higher level.

As it is shown in figure 3-B, our proposed higher-level kinematic strategy consists of three stages and systematically combines the three previously mentioned motor mechanisms. In the first stage, head remains fixed while the eye is moving in the head. In the second stage, head rotates, moving eye with itself such that eye-in-head position remains unchanged; this rotation foveates the target. In the third stage, head rotates to its definite position, while eye rotates in head to compensates for head movement and keeps the target foveated (vestibulo-ocular reflex). Table 1 shows the orientations of the eye, head and gaze after any of the three stages of the model.

	$\vec{e}$	$\vec{h}$	$\vec{g}$
Initial Condition	$\boldsymbol{E_i} \times \vec{e}_r$	$\boldsymbol{H_i} \times \vec{h}_r$	$\boldsymbol{H_i} \times \boldsymbol{E_i} \times \vec{g}_r$
After 1st Stage	$\boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{e}_r$	$\boldsymbol{H_i} \times \vec{h}_r$	$\boldsymbol{H_i} \times \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{g}_r$
After 2nd Stage	$\boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{e}_r$	$\boldsymbol{Rh} \times \boldsymbol{H_i} \times \vec{h}_r$	$\boldsymbol{Rh} \times \boldsymbol{H_i} \times \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{g}_r$
Desired Condition	$\boldsymbol{Rv} \times \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{e}_r$	$\boldsymbol{Rw} \times \boldsymbol{Rh} \times \boldsymbol{H_i} \times \vec{h}_r$	$\boldsymbol{Rw} \times \boldsymbol{Rh} \times \boldsymbol{H_i} \times \boldsymbol{Rv} \times \boldsymbol{Re} \times \boldsymbol{E_i} \times \vec{g}_r$

Table 2: mathematical description of eye, head and gaze orientations at different stages

So, desired orientations can be written as a function of initial orientations and the rotations:

$$\boldsymbol{E_d} = \boldsymbol{Rv} \times \boldsymbol{Re} \times \boldsymbol{E_i} \tag{25}$$

$$\boldsymbol{H_d} = \boldsymbol{Rw} \times \boldsymbol{Rh} \times \boldsymbol{H_i} \tag{26}$$

$$\boldsymbol{G_d} = \boldsymbol{Rw} \times \boldsymbol{Rh} \times \boldsymbol{H_i} \times \boldsymbol{Rv} \times \boldsymbol{Re} \times \boldsymbol{E_i} \tag{27}$$

4.3.5 Solving the Static Model

Dependence of Desired Head Position on Desired Gaze Position

When the desired target appears in the visual field, the main signal for planning the gaze-shift and the main known input of our model is constructed in the form of the 2-D desired

angular gaze position $\vec{c}_d$. We define the parameters α, β to determine how much the head would move, in horizontal and vertical directions respectively, relative to initial head position. Setting α, β to zero, the model is reduced to a model of head-fixed gaze-shift. Model parameters α, β determine how the 2-D desired angular head position $\vec{b}_d$ would be derived from $\vec{c}_d$ and the initial conditions:

$$\vec{b}_d = [b_i(1) + \alpha \times (c_d(1) - b_i(1)) \; ; \; b_i(2) + \beta \times (c_d(2) - b_i(2))] \tag{28}$$

Where $0 < \alpha < 1$ and $0 < \beta < 1$. Desired gaze and head vectors, $\vec{g}_d$ and $\vec{h}_d$, can then be derived from $\vec{c}_d$ and $\vec{b}_d$ based on equations (7) and (8).

Fick Constraint for Head Orientation

Fick system represents a general rotation as successive rotations with magnitudes θ, φ, ψ about local vertical, horizontal and torsional axes respectively. Rotation matrix in Fick system is:

$$\begin{bmatrix} \cos(\varphi)\cos(\theta) & -\cos(\varphi)\sin(\theta) & \sin(\varphi) \\ \cos(\psi)\sin(\theta) + \sin(\psi)\sin(\varphi)\cos(\theta) & \cos(\psi)\cos(\theta) - \sin(\psi)\sin(\varphi)\sin(\theta) & -\sin(\psi)\cos(\varphi) \\ \sin(\psi)\sin(\theta) - \cos(\psi)\sin(\varphi)\cos(\theta) & \sin(\psi)\cos(\theta) + \cos(\psi)\sin(\varphi)\sin(\theta) & \cos(\psi)\cos(\varphi) \end{bmatrix} \tag{29}$$

It has been shown that after a natural head-free gaze-shift, desired head orientation obeys the Fick constraint. This constraint states that if one represents $\boldsymbol{H_d}$ in the Fick system, then the torsional component of this representation is zero:

$$\boldsymbol{H_d} = \begin{bmatrix} \cos(\rho_{\boldsymbol{H_d}})\cos(\theta_{\boldsymbol{H_d}}) & -\cos(\rho_{\boldsymbol{H_d}})\sin(\theta_{\boldsymbol{H_d}}) & \sin(\rho_{\boldsymbol{H_d}}) \\ \sin(\theta_{\boldsymbol{H_d}}) & \cos(\rho_{\boldsymbol{H_d}})\cos(\theta_{\boldsymbol{H_d}}) & 0 \\ -\sin(\rho_{\boldsymbol{H_d}})\cos(\theta_{\boldsymbol{H_d}}) & \sin(\rho_{\boldsymbol{H_d}})\sin(\theta_{\boldsymbol{H_d}}) & \cos(\rho_{\boldsymbol{H_d}}) \end{bmatrix} \tag{30}$$

Knowing $\vec{\mathrm{h}}_\mathrm{d}$, Fick angles of desired head orientation, $\theta_{\boldsymbol{H_d}}, \rho_{\boldsymbol{H_d}}$, can be derived based on general relation (2) and the equation (30):

$$\theta_{\boldsymbol{H_d}} = \sin^{-1}(h_d(2)) \tag{31}$$

$$\rho_{\boldsymbol{H_d}} = \sin^{-1}\left(-\frac{h_d(3)}{\cos(\sin^{-1}(h_d(2)))}\right) \tag{32}$$

So, desired head orientation $\boldsymbol{H_d}$ would now become known to us.

Uniqueness of Head Rotation Command

From observations of the behavior in head-unrestrained experiments, it has been seen that only one head rotation command is implemented during one planned gaze-shift. However, two distinct measures of the head movement have been defined: the total movement of the head from start to finish and the amount that the head movement contributed to the accomplishment of the gaze-shift, often referred to as the head contribution (Bizzi et al., 1972, Morasso et al., 1973). So, in our model structure, we assume that the head rotations in the 1st and 2nd stages of our model are just two successive parts of one head rotation $\boldsymbol{Rt}$:

$$\boldsymbol{Rt} = \boldsymbol{Rw} \times \boldsymbol{Rh} \tag{33}$$

This means that $\boldsymbol{Rw}$ and $\boldsymbol{Rt}$ have the same axis of rotation:

$$\vec{u}_{\boldsymbol{Rt}} = \vec{u}_{\boldsymbol{Rw}} = \vec{u}_{\boldsymbol{Rh}} \tag{34}$$

And rotation magnitudes of $\boldsymbol{Rh}$ and $\boldsymbol{Rw}$ are complementary fractions of $\tau_{\boldsymbol{Rt}}$:

$$\tau_{\boldsymbol{Rh}} = \delta \times \tau_{\boldsymbol{Rt}} \tag{35}$$

$$\tau_{\boldsymbol{Rw}} = (1 - \delta) \times \tau_{\boldsymbol{Rt}} \tag{36}$$

Where $0 < \delta < 1$ and δ is a model parameter which could depend on different factors, most importantly the total head rotation. After finding $\boldsymbol{H_d}$ from equations (30-32), we can derive $\boldsymbol{Rt}$ based on equations (26) and (33):

$$\boldsymbol{Rt} = \boldsymbol{H_d} \times \boldsymbol{H_i}^{-1} \tag{37}$$

$\boldsymbol{Rh}$ and $\boldsymbol{Rw}$ will be found as we know their axis and magnitude of rotation.

Listings' Law for Eye Orientation

$\mathbf{H_d}$ and $\vec{g}_d$ being known, we can find $\vec{e}_d$ from:

$$\vec{e}_d \;=\; \boldsymbol{H_i}^{-1} \times \vec{g}_d \tag{38}$$

Based on Listing's law, if one represents eye-in-head orientation by the classical magnitude/axis convention, then the axis of rotation would always be in the Listings plane (LP). LP is a plane fixed to the head and rotating with it. LP is orthogonal to the straight ahead sight/gaze axis. According to this constraint, the third component of the unit vector, which denotes the axis of rotation for eye-in-head orientation matrix, is zero. For the desired eye-in-head orientation:

$$\vec{u}_{\boldsymbol{E_d}} = [u_{\boldsymbol{E_d}}(1); u_{\boldsymbol{E_d}}(2); 0] \tag{39}$$

$$E_d = \begin{bmatrix} cos(\tau_{E_d}) + u_{\boldsymbol{E_d}}(1)^2 \times (1 - cos(\tau_{E_d})) & u_{\boldsymbol{E_d}}(1) \times u_{\boldsymbol{E_d}}(2) \times \left(1 - cos(\tau_{E_d})\right) & +u_{\boldsymbol{E_d}}(2) \times sin(\tau_{E_d}) \\ u_{\boldsymbol{E_d}}(1) \times u_{\boldsymbol{E_d}}(2) \times \left(1 - cos(\tau_{E_d})\right) & cos(\tau_{E_d}) + u_{\boldsymbol{E_d}}(2)^2 \times (1 - cos(\tau_{E_d})) & -u_{\boldsymbol{E_d}}(1) \times sin(\tau_{E_d}) \\ -u_{\boldsymbol{E_d}}(2) \times sin(\tau_{E_d}) & u_{\boldsymbol{E_d}}(1) \times sin(\tau_{E_d}) & cos(\tau_{E_d}) \end{bmatrix} \tag{40}$$

Substituting (40) into (1) and knowing $\vec{e}_{\mathrm{d}}$ from (38), we can solve the system of equations for $u_{\boldsymbol{E_d}}(1)$ and $u_{\boldsymbol{E_d}}(2)$ and τ_{E_d}:

$$\tau_{\boldsymbol{E_d}} = \cos^{-1}(e_d(3)) \tag{41}$$

$$u_{\boldsymbol{E_d}}(1) = {-e_d(2)}\Big/{sin(\tau_{\boldsymbol{E_d}})} \tag{42}$$

$$u_{\boldsymbol{E_d}}(2) = {e_d(1)}\Big/{\sin(\tau_{\boldsymbol{E_d}})} \tag{43}$$

So, from (40) we have $\boldsymbol{E_d}$. Let's define rotation matrix $\boldsymbol{Ra}$ as:

$$\boldsymbol{Ra} \;=\; \boldsymbol{Re} \times \boldsymbol{Rv} \tag{44}$$

Knowing $\boldsymbol{E_i}$ and $\boldsymbol{E_d}$, we can derive $\boldsymbol{Ra}$ from (25):

$$\boldsymbol{Ra} \;=\; \boldsymbol{E_d} \times \boldsymbol{E_i}^{-1} \tag{45}$$

Gaze Stability during VOR

We are assuming that gaze direction does not change during the third stage of the model and by the execution of $\boldsymbol{Rw}$ and $\boldsymbol{Rv}$. Then, by looking at the table 1, we have:

$$\boldsymbol{Rh \times H_i \times Re \times E_i = Rw \times Rh \times H_i \times Rv \times Re \times E_i} \quad (46)$$

From (46), we can derive $\boldsymbol{Rv}$**:**

$$\boldsymbol{Rv = H_i^{-1} \times Rh^{-1} \times Rw^{-1} \times Rh \times H_i} \quad (47)$$

Knowing $\boldsymbol{Ra}$ and $\boldsymbol{Rv}$, $\boldsymbol{Re}$ can be derived from (44):

$$\boldsymbol{Re = Rv^{-1} \times Ra} \quad (48)$$

Therefore, all the unknown parameters of the model have been derived from the governing equations of the model considering the assumptions and constraints.

4.3.6 Simulation of full Movement Trajectories

The model described above was designed to simulate the key kinematic events in the gaze shift illustrated in Figure 3-B. For simulation purposes, this was sufficient to show initial and final eye (saccade and VOR) and head movement positions. A complete dynamic model of the system would require neural and mechanical elements downstream from the model in Figure 1, and goes beyond the goals and scope of the current paper. However, for some of the simulations shown below it was desirable esthetically or scientifically to show intermediate points along the entire trajectory. In brief, to do this we assumed constancy of the axis of rotation for all eye and head motions except VOR (whose axis of rotation is determined online from the online spatial orientation of head). We then discretized the magnitude of rotation based on specific growth functions in a time-frame illustrated in figure 3-C. The 3-D constraints in our model were applied on initial and final eye/head orientation and we do not analyze velocity or acceleration in this paper, so, the details of these growth functions have no bearing on any of the questions asked here.

4.4 Results & Discussion

Here we test the model by comparing its simulated output to previously reported or expected performance of the real system in several different tasks. Unless otherwise stated, the model parameters are set to $\alpha = \beta = \delta = 0.5$.

4.4.1 Gaze Accuracy and the 3-D Reference Frame Transformations

It has been shown both with saccade simulations (Crawford and Guitton, 1997c) and real saccade data (Klier and Crawford, 1998) that retinal error only corresponds directly to the gaze movement vector for saccades directed toward, across, or away from Listing's primary eye position. For all other saccades, retinal error needs to be mapped onto different saccade vectors as a function of initial eye orientation. This is simply a function of the geometry of the system; it cannot be any other way. However, failure to properly account for this, in our model (or the real gaze control system), would result in saccade errors that increase with the position component and length of retinal error (which did not occur). This has not been measured behaviorally with head-unrestrained gaze shifts, but the predicted errors here would be so large (up to 90º) that it is obvious that the system accounts for this. Moreover, the converse has been shown with simulations and experimentally: a single retinal vector evoked from stimulation of the brain (e.g. in superior colliculus) results in very different eye-head gaze trajectories as a function of initial eye orientation (Klier et al., 2001, Martinez-Trujillo et al., 2004). Tweed's (1997) model supports this behavior, and was used as the basis for the latter simulations.

We have simulated this behavior with our model in figure 4. Here, the model generated rightward gaze-shifts from different vertical positions but the same horizontal components (○), either with a fixed rightward gaze trajectory toward the symmetric target on the opposite side (left column) or with a fixed rightward retinal error input (right column). The intersection point of gaze on a forward-facing target screen is shown in the top row (with end points shown as ×), the instantaneous points of stimulation of the corresponding positions on the retina (initial retinal error being the main model input) are shown in the middle row, and the resulting angular gaze trajectories (the model output) are shown in the bottom row. The trajectories in the upper and bottom rows are very similar, starting from the left and proceeding right. The trajectories in the middle row proceed in the opposite direction because the desired targets start to the right in retinal coordinates (dispensing with the optical inversion) and then proceed to the left as they converge toward (0,0), i.e. the retinal coordinates of the fovea. This indicates the accuracy of the model in bringing the image of the desired target to the fovea.

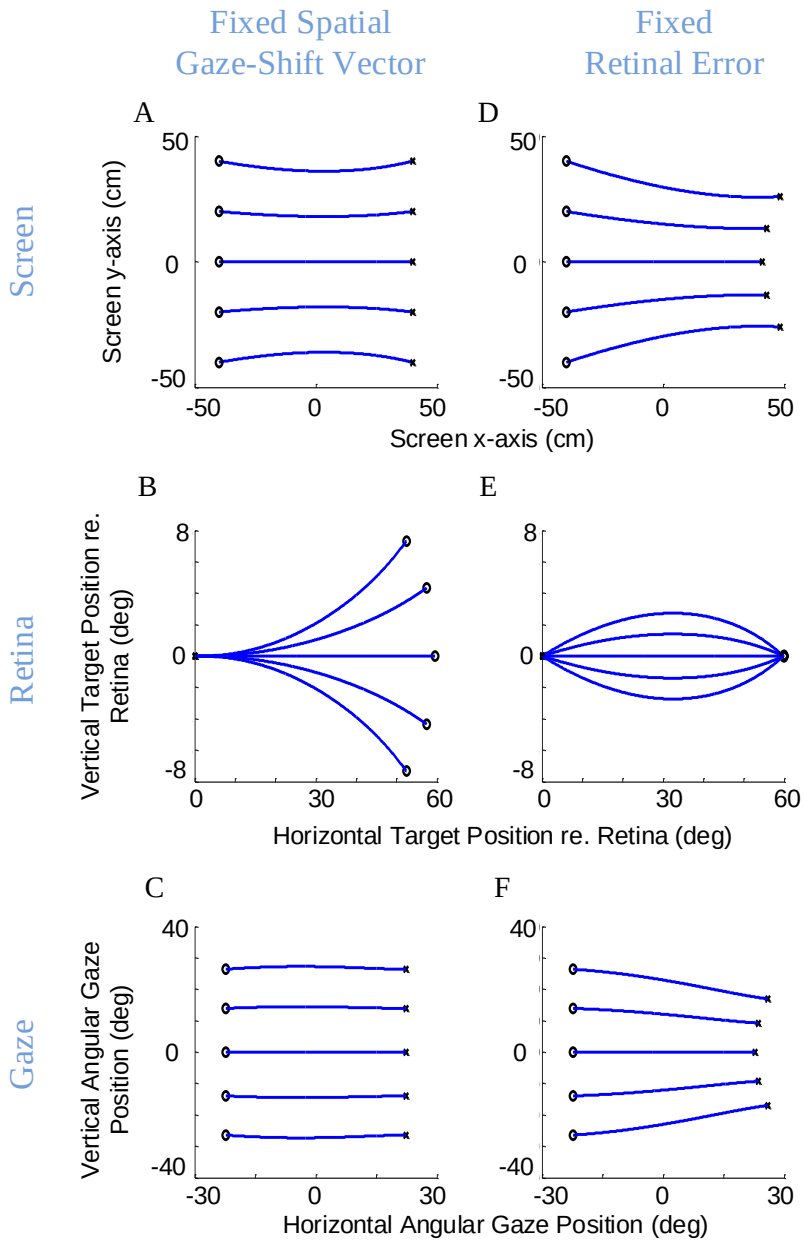

Figure 4: **Gaze Accuracy and the 3-D Reference Frame Transformations for Gaze-Shifts**. Rightward gaze-shifts are simulated from five different vertical altitudes with either a fixed symmetric horizontal gaze-shift, -40cm left to 40cm right, on a flat target screen (left column), or from the same initial positions with a fixed retinal error of 60 degrees right (right column). First row shows the initial and desired target positions on the screen and the development of gaze direction on the screen during the gaze-shift. Second row shows the development of the target position in retinal coordinates during gaze-shift. Third row shows the development of the 2-D angular gaze position during gaze-shift. For both conditions, the model parameters are set to $\alpha = ß = \delta = 0.5$. Circles show initial target locations while stars show the desired positions of target. Note that in B even though the targets are due right in spatial coordinates, they have variable vertical components in retinal coordinates, whereas conversely retinal errors in E start and end at the same positions, and correspond to different gaze trajectories.

More importantly, these simulations illustrate the non-trival relationship between retinal error vectors and gaze shift direction, and the ability of our model to handle this. As the left column shows, when the target is due right of initial gaze position (4-A), this corresponds to non-horizontal retinal errors (4-B) as a non-linear function of initial vertical position, but the model correctly converts this into rightward gaze shifts (4-C). Conversely, the right column shows that a rightward retinal error (4-E) corresponds to different directions of target position releative to initial position (4-D), but again the correct movement trajectory is generated (4-F). We obtained analogous results for every combination of retinal error and position that we tested. There can be no linear trivial mapping between the retina and motor output, unless one models the pulling actions of the eye and neck muscles into retinal coordinates and aligns the centres of rotation of the eyes and head, which is not realistic. Thus, the model must (and does) perform an internal reference frame transformation, based on its retinal inputs and its eye / head orientation inputs.

4.4.2 Eye, Head and Gaze Orientations and their Constraints

Donders' law, as originally stated, suggested that the eye should only attain one torsional orientation for each gaze direction, irrespective of the path taken to acquire that position. This rule has since been applied and elaborated to a number of situations and more specific rules. Behavioral data from 3-D head-fixed and head-free tasks (Glenn and Vilis, 1992, Radeau, 1994, Crawford et al., 1999b) have shown that the 1) orientation of eye relative to head at the end of the gaze-shift lies in the Listing's plane and has zero torsional component, 2) the final orientation of head relative to shoulder obeys the Fick law, i.e. the torsional component of head orientation in Fick system is zero, and 3), the orientation of the eye-in-space during gaze fixations also adheres to a form of Donder's law similar to the Fick rule.

Importantly, in our model, the Listings and Fick constraints on final eye and head orientation were directly implemented, whereas in our model torsion of the eye in space was an emergent property of the above constraints. What would this look like? The final positions of gaze-shifts of various amplitudes and directions are simulated in figure 5 for the eye-in-head (left column), head-in-space (middle column), and eye-in-space (right

column), where the 1st row shows the 2-D components of this range and the 3rd row shows horizontal position plotted as a function of torsional position. As one can see in figure 5-D, irrespective of the magnitude or direction of eye or head rotations during gaze-shifts, this kinematic model always produces a final eye-in-head orientation that obeys Listing's law, forming a flat range of zero torsional positions. In contrast, Fick constraint manifests as a bow-tie shape of the distribution of head orientations in horizontal-torsional rotation plane. As one can see in figure 5-E, all final head positions, irrespective of the magnitude or direction of head rotation, obey the Fick law for head orientation. A similar, but less pronounced, Fick-like twist in the range of final orientation is seen for the eye in space (Figure 5-F). In other words, in our model the Fick range of eye orientation in space was an emergent property of the eye and head constraints implemented in our model.

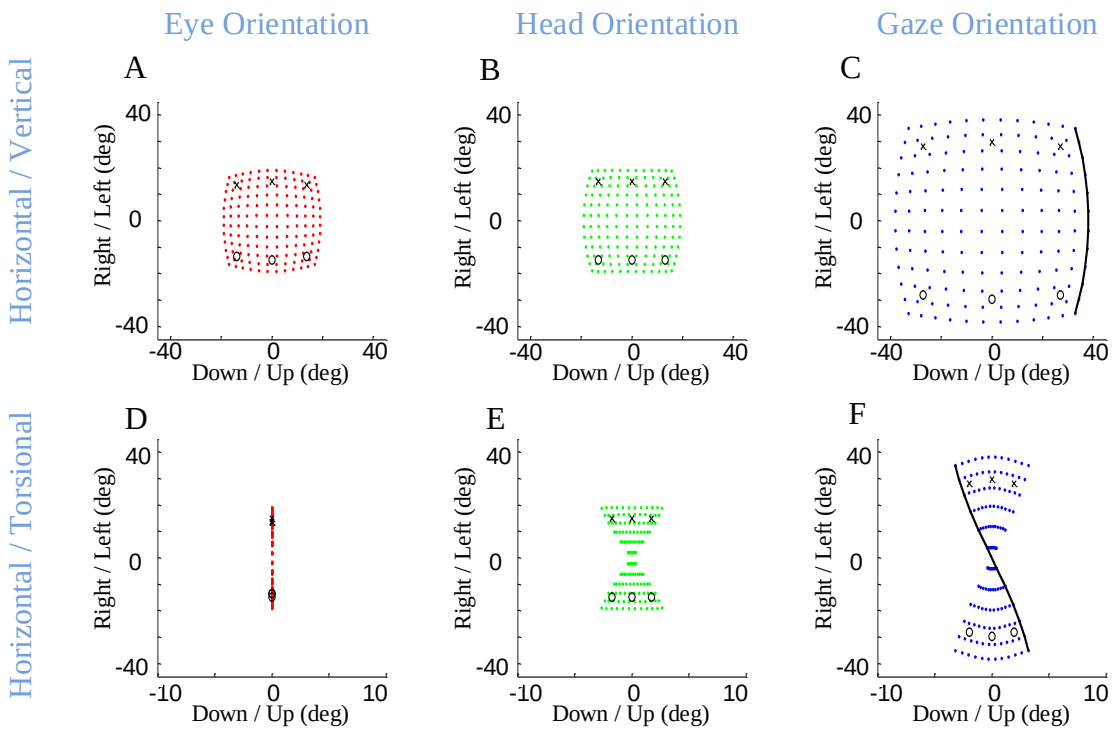

Figure 5: **Distributions of Head, Eye and Gaze Orientations for Equal Contributions of Eye and Head Rotations to Horizontal and Vertical Directions**. Model simulations producing gaze-shifts from the central fixation point (reference condition) to a uniform distribution of targets on the screen in range (-40,40) degrees horizontal and (-40,40) degrees vertical. The first, second and third columns respectively show eye-in-head (red), head-in-space (green) and eye-in-space orientations after the gaze-shift. First row illustrates the horizontal (right/ left) against the vertical (up/down) components while the third row shows the horizontal (right-left) against the torsional (CW/CCW) components. The parameters of the model are set to $\alpha = ß = \delta = 0.5$. The black curve shows gaze orientations for targets aligned horizontally on top of the screen.

Tweed (1997) was also able to simulate similar behaviors, but in that case it was assumed that eye-in-space orientation was explicitly controlled and other 3-D parameters were derived from this. However, there are maybe some potential differences between the predictions of our models. First, our model seems to be more consistent with the consistent observation that eye-in-space torsion is more variable than eye or head torsion (Glenn and Vilis, 1992, Radeau, 1994, Crawford et al., 1999b). We could simulate this by summating independent random noise in eye and head torsion, but the result would be trivial. In contrast, Tweed's model could predict a tighter constraint on eye-in-space torsion (particularly if errors in the decomposition of 3-D gaze commands produced anti-correlated noise in eye and head torsion).

4.4.3 Development of the Eye, Head and Gaze Orientations during Gaze-Shift

The previous section only described end point kinematics of the entire eye-head gaze shift. The 3-D trajectories of the eye and head during motion, and their relationship to the end-point constraints, are potentially much more complex. It is generally agreed that Listing's law is obeyed during head-restrained saccades (Ferman et al., 1987b, Tweed and Vilis, 1990), although small torsional 'blips' near the ends of the trajectories have been scrutinized to test the role of eye mechanics in implementing the position-dependent 'half angle rule' that describes 3-D eye velocities for Listing's law (Straumann et al., 1995, Straumann et al., 1996). We have assumed that these rules are perfectly implemented downstream from the output of our model so our model cannot predict any such 'blips'. However, eye trajectories become much more complicated in the head-unrestrained situation because saccades must be coordinated with the VOR, which does not obey Listing's law, resulting in large transient deviations of eye position from Listing's plane (Crawford and Vilis, 1991, Tweed et al., 1998, Crawford et al., 1999b, Klier et al., 2003). Likewise, during rapid gaze shifts in monkeys the head appears to deviate from the static Fick constraint when it takes the shortest path between two points on the curved Fick range (Crawford et al., 1999b). These saccade/VOR behaviors have been considered in a previous modeling study (Tweed, 1997), but not the above-mentioned head behavior.

Here, we consider the ability of our model to simulate these behaviors, based on its static implementation of the Listing and Fick rules, and the simple discretization of trajectories

described in section 2. Figure 6 shows example eye, head, and gaze trajectories between three initial (○) and final (×) gaze positions (corresponding to the same symbols / positions shown in Figure 5). Figure 6 thus shows the development of gaze-shift, in different rotation planes, between two groups of vertically aligned targets on the screen. Likewise, figure 7 illustrates the temporal development of the horizontal (upper row), vertical (middle row) and torsional (bottom row) components of eye (left column), head (middle column) and gaze (right column) orientations during the same set of gaze shifts as shown in figure 6.

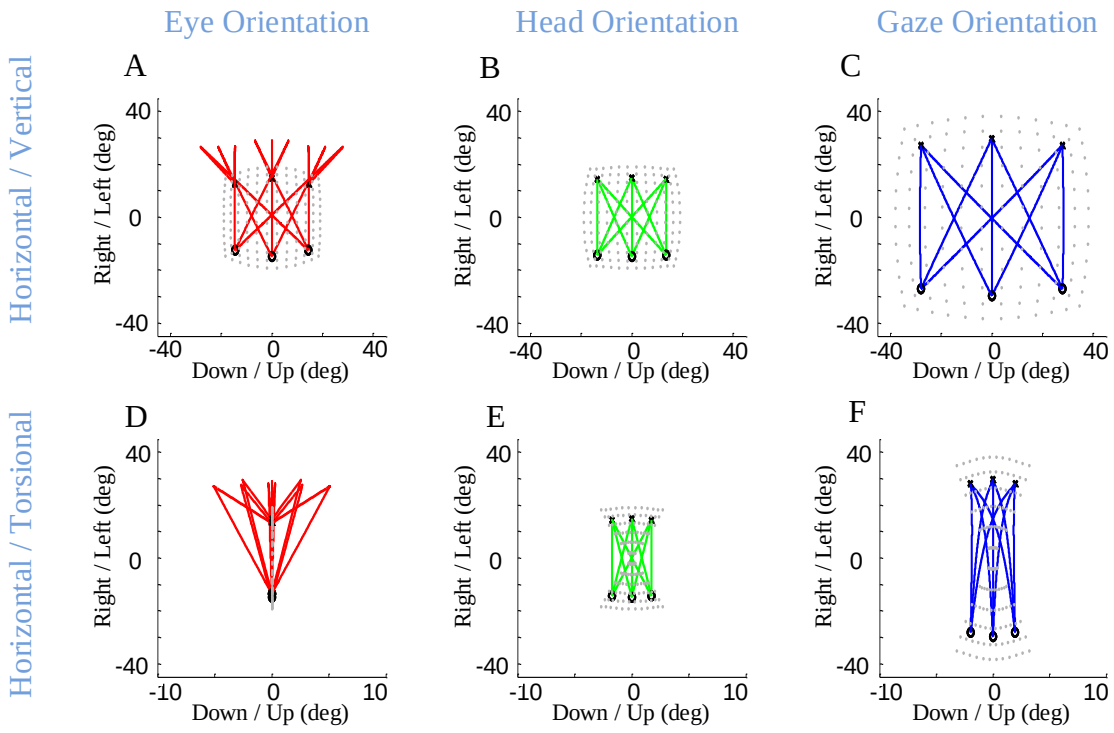

Figure 6: **Spatial Path of the Development of Eye, Head and Gaze Orientations during Gaze-Shift.** Three example gaze-shifts have been planned from three targets, vertically aligned at -40 cm on the screen, to another three targets, vertically aligned at 40 cm. The locations of eye, head and gaze in initial condition are shown by circles while their locations in desired condition are shown by crosses. First and second rows show the temporal development of eye, head and gaze in vertical-horizontal and torsional-horizontal planes respectively.

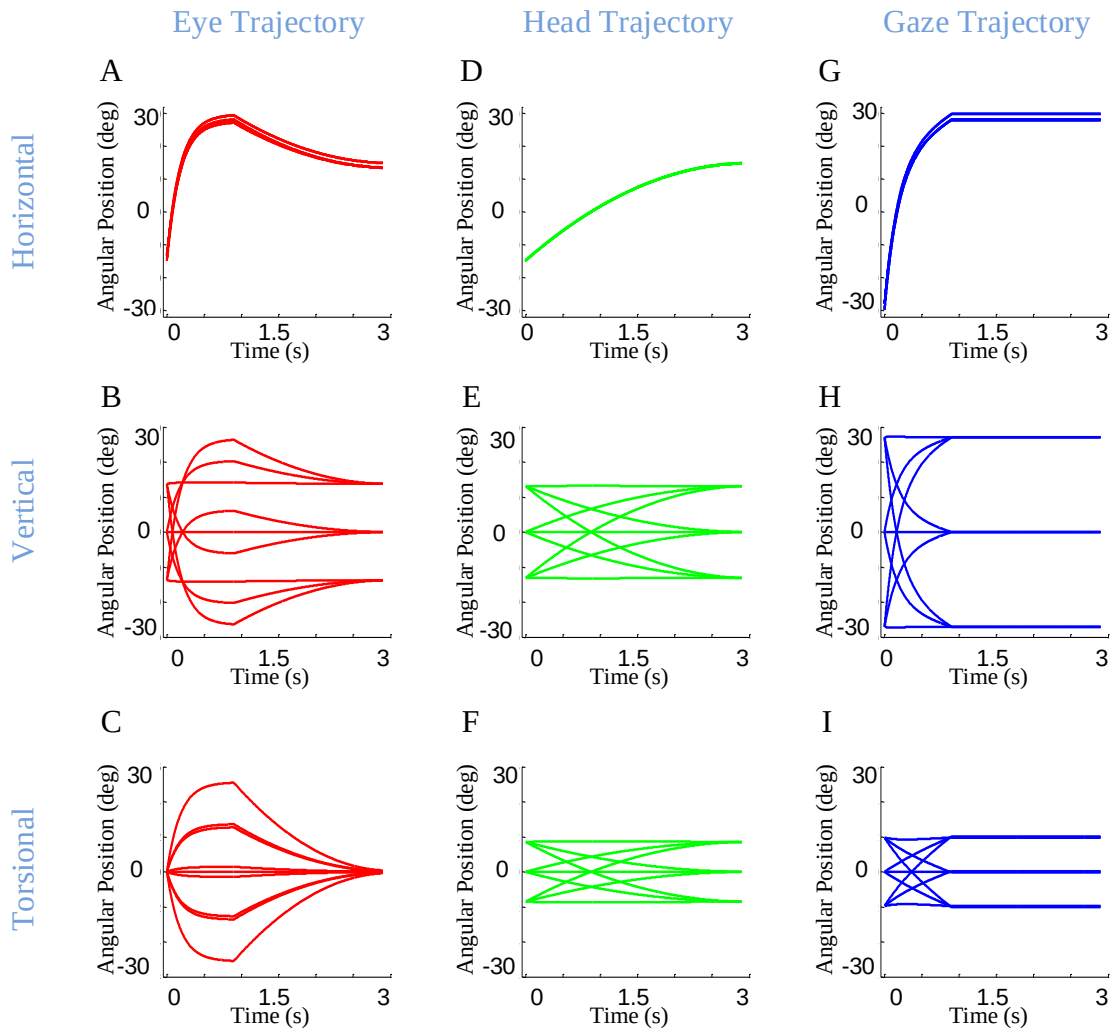

Figure 7: **Temporal Pattern of Development of Eye, Head and Gaze Orientations during Gaze-Shift.** For the same nine gaze-shifts, between two groups of vertically aligned targets, we have shown the development of the orientations. 1st, 2nd, and 3rd columns show orientations of eye, head and gaze respectively. 1st, 2nd and 3rd rows describe the development of horizontal, vertical and torsional components of orientations respectively.

First, we consider the eye-in head behavior. In real time, the VOR is evoked through vestibular stimulation after the saccade, but in our model (and we propose in real physiology) the brain implicitly predicts the VOR from intended head movement signals in order to program the right amount of torsion and also bring the eye onto the correct final 2-D orientation (Crawford and Guitton, 1997a, Misslisch et al., 1998). This is illustrated in the left columns of Figure 6 and Figure 7. Eye orientation relative to head goes out of its range during saccade and comes back to the planned configuration by VOR (Fig 6-A). Particularly, the eye-in-head torsion starts at the Listings' plane, deviates from the LP during the saccade and gets back into the LP by the VOR (Fig 6-D). The reasons for this are more clearly illustrated in figure 7. Here, one can see that the gaze-shift is implemented in two time phases: 1- Eye undergoes a saccade, head contributes to gaze, and gaze is placed on the target. 2- Head undergoes its second-stage movement (cancelled out by the VOR), the eye is driven by the VOR, and gaze is stabilized. The eye torsion (Figure 7-C) starts at zero which indicates that initial eye orientation obeys the Listing's law. Thus, torsions in these two phases neutralize each other such that the torsion of the final eye orientation is zero in Listings' plane coordinates. Similar principles hold for horizontal and vertical eye position (Figure 7-A, 7-B), except that these saccade components are larger than the corresponding VOR components. This replicates the behaviors observed in monkey and human gaze shifts (Crawford and Vilis, 1991, Tweed et al., 1998, Crawford et al., 1999b, Klier et al., 2003).

In our model, the head's Fick constraint is only explicitly specified at its initial and final positions, and the head is moved uniformly through the gaze shift by a single rotation command. As a result, in our simulations, the head starts and ends in the Fick range, moves smoothly between these positions, and often violates the Fick constraint during the movement (Figure 6-E, Figure 7-F). The deviations from Fick are made clear by comparing figures 6-E and 5-E, which has been imposed in gray beneath 6-E for easy reference. If the head always obeyed the Fick constraint during gaze-shifts, it would take a path passing through the bow-tie shape. Instead, the head takes an almost direct path between the two Fick-obeying points. This is most clear in the head movements between corners with similar torsion (e.g. the two left-side corners and two right side corners in Fig 6-E), where the head completely leaves the normal Fick range. This replicates the experimental

observations in the monkey (Crawford et al., 1999b). However, more experiments are required to know if the head always follows the same strategy.

Finally, note again that in our model, gaze (eye orientation in space) torsion is not explicitly controlled during the trajectory either, but is rather is an emergent property (roughly the simultaneous sum) of eye and head torsion during the gaze shift. Thus, not surprisingly, gaze torsion also deviates from its normal qausi-Fick range during the gaze shift (Fig. 7F).

4.4.4 Eye-Head Coordination Strategies Influence Eye-in-Space Orientation

During visual fixations, the entire 3-D range of eye orientation is important because this determines the orientation of the retina relative to the visual world (Ronsse et al., 2007). However, this topic (eye orientation in space) has received surprisingly little attention compared to 2-D gaze direction. Our physiologically-inspired model assumes that eye-in-space torsion is an emergent property of separate constraints on eye and head torsion. As we shall see, this gives rise to the possibility that eye-head coordination strategies could interact with these constraints to produce different ranges of eye orientation in space. In this section we consider several possible, experimentally testable situations where this could occur.

It has been shown in many experiments (and is also intuitively obvious from personal experience) that the amount that the head rotates for a constant gaze-shift changes depending on many factors, including initial head orientation (Guitton and Volle, 1987), visual range (Crawford and Guitton, 1997a), behavioral context (Land, 1992), expected future gaze targets (Monteon et al., 2012), and inter-subject differences. In order to reflect this variability, we have defined two variables α & β (changing in range $[0, 1]$) which respectively determine the horizontal and vertical angular positions on which head falls after the gaze-shift. This allowed us to explore the kinematic consequences of 1) utilizing different overall eye vs. head contributions to gaze-shift, and 2) differential vertical vs. horizontal contributes of the head to gaze-shift.

Infinitesimal values of α & β correspond to nearly head-fixed saccades (Fig 8, top row), reflecting situations such as watching television and reading (Proudlock et al., 2003). Here, eye orientation occupies almost the same area as gaze (Fig 8-A vs. C) while head

orientation is limited to a very small area (Fig 8-B). In this condition, gaze orientation comes very close to following Listing's law (Fig 8-C). In contrast, large values of α & β (Fig 8, bottom row) were used to simulate the situation where final head orientations occupied almost the same area as gaze distribution, and eye-in-head orientation returns to a central range near primary position after the VOR. This emulates behavioral situations such as driving a car (Land, 1992) and certain experiments in which subjects were required to rotate their head more (Ceylan et al., 2000). Here, the head's greater contribution to gaze orientation (while maintaining final eye-in head torsion at zero) results in a Fick-like range of eye-in-space orientations identical to that of the head (Fig 5-F)., and thus more 'twisted' than observed when the eye and contribute equally.

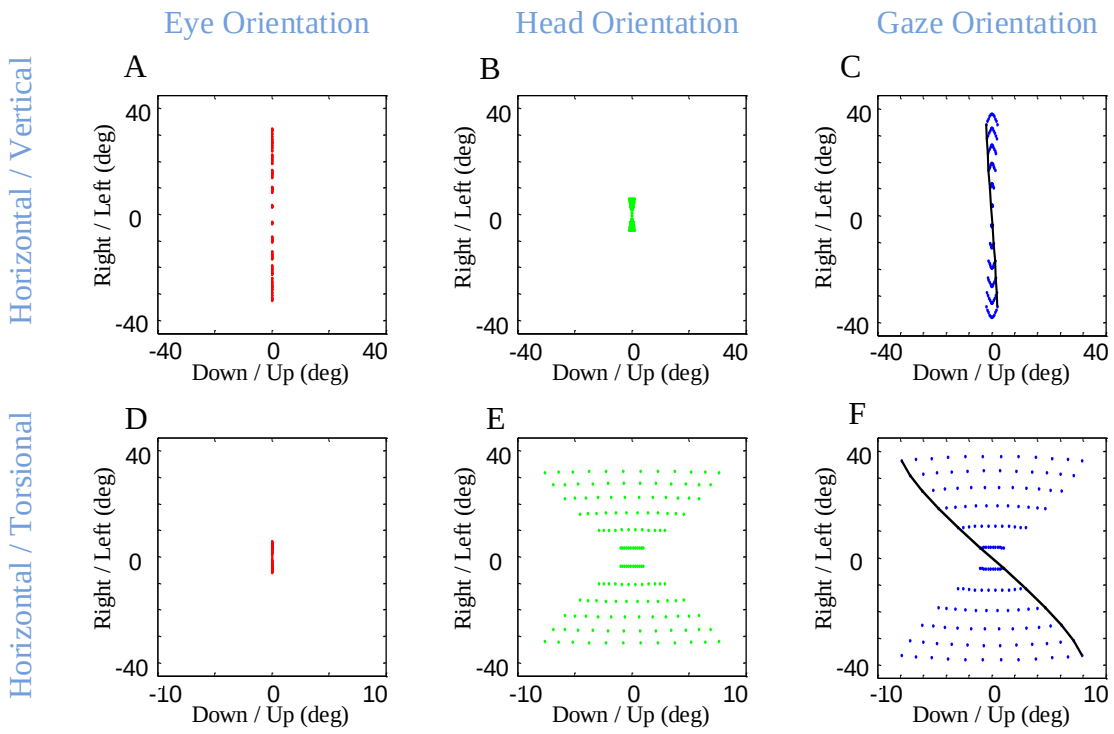

Figure 8: **Distributions of Head, Eye and Gaze Orientations for Two Extreme Cases of Almost Only Eye Contribution (Head-Fixed Saccade) and Almost Only Head Contribution.** We have made the model to plan gaze-shifts from the central fixation point (reference condition) to a uniform distribution of targets on the screen in range (-40,40) degrees horizontal and (-40,40) degrees vertical. Eye-in-head (first column in red), head-in-space (second column in green) and eye-in-space (third column in blue) orientations are illustrated. Only the horizontal (right-left) against the torsional (CW/CCW) diagrams are included in this figure. The parameters of the model for the first row is set to α = ß = 0.15 and δ = 0.5 while for the second row they are set to be α = ß = 0.85 and δ = 0.5. The black curve shows gaze orientations for targets aligned horizontally on top of the screen.

Note that the latter simulations assumed that constraints on eye and head orientation are not influenced by these different eye head coordination strategies. To our knowledge, this has not been directly tested for the 'eye-only' situation, but, experimental studies that increased the amount of head orientation to equal gaze orientation (by training subjects to look through a head-fixed 'pinhole' or point a head-fixed light toward the target) caused the head to develop a more Listing-like strategy (Crawford et al., 1999b, Ceylan et al., 2000) and thus producing a less twisted eye-in-space range our simulation. This could be simulated here by replacing our head's Fick constraint with a Listing's law constraint as used in the eye pathway. The more important point is that in the Ceylan et al. (2000) study concluded that these head constraints are purely motor, whereas the current analysis suggests that their result might have been related to orientation of the eye in space and its implications for vision (see section 3.5). If so, then the brain would have to be aware of the interactions between eye-head coordination and 3-D orientation constraints, and alter the latter accordingly to achieve the right position range.

Another interaction between eye-head coordination and orientation constraints is perhaps more surprising, and yet inevitable if the assumptions behind our model are correct. It has been experimentally observed that the contribution of the head to the gaze-shift can be different in horizontal and vertical directions, usually providing more horizontal contribution (Freedman and Sparks, 1997, Crawford et al., 1999b). Figure 9 shows the ability of the model to plan such distinct gaze-shifts, and uses these simulations to illustrate how relative vertical-horizontal contributions of the head to gaze shifts could have a profound influence on orientation of the eye in space.

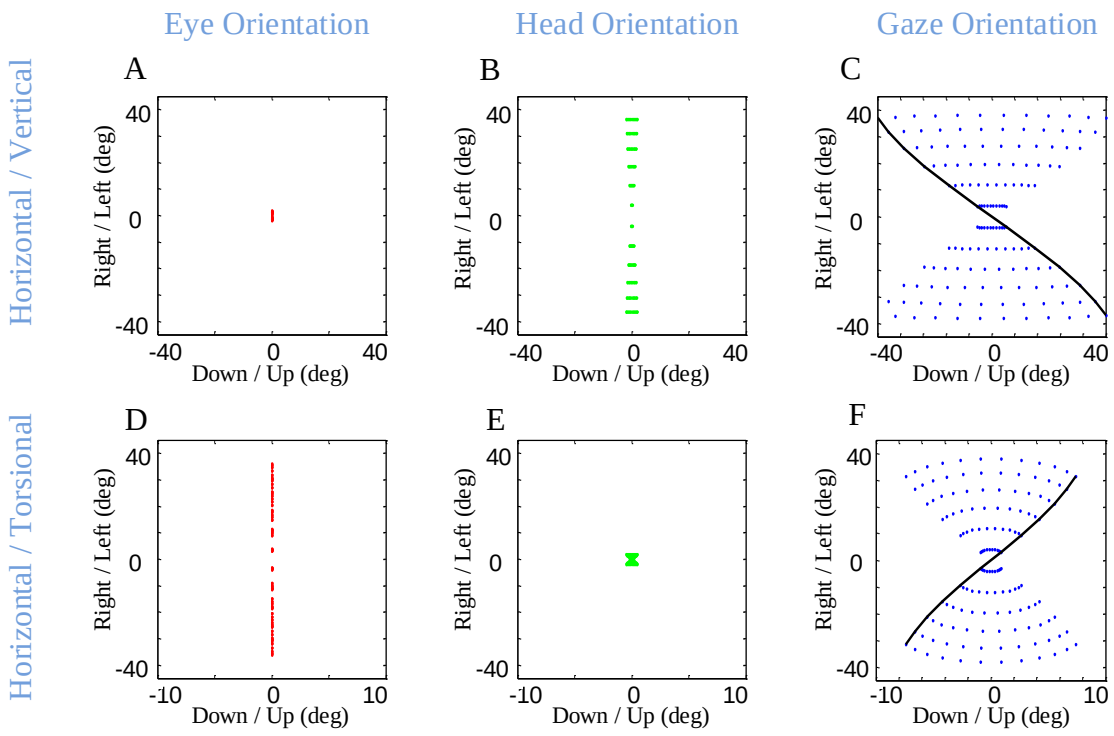

Figure 9: **Distributions of Head, Eye and Gaze Orientations for Two Extreme Cases of Almost Only Head Contribution to Horizontal Gaze-Shift or Almost Only Head Contribution to Vertical Gaze-Shift.** We have made the model to plan gaze-shifts from the central fixation point (reference condition) to a uniform distribution of targets on the screen in range (-40,40) degrees horizontal and (-40,40) degrees vertical. Eye-in-head (first column in red), head-in-space (second column in green) and eye-in-space (third column in blue) orientations are illustrated. The horizontal (right-left) against the torsional (CW/CCW) diagrams are only included in this figure. The parameters of the model for the first row is set to $\alpha = 0.05$, $ß = 0.95$, and $\delta = 0.5$ while for the second row they are set to $\alpha = 0.95$, $ß = 0.05$, and $\delta = 0.5$. The black curve shows gaze orientations for targets aligned horizontally on top of the screen.

In the first row of figure 9, the eye (A) contributes mainly to vertical component (not shown) and the head (B) is mainly contributing to the horizontal component of the gaze-shift. This essentially reduces eye and head orientation each to rotation about two fixed axes and a one-dimensional range, but results in a strong 'Fick-like' twist in the eye-in-space orientation range (C), even stronger than in our default simulations (Fig 5-F). This is because here we have essentially turned the system into a true Fick Gimbal, where the head rotates about a body-fixed vertical axis and the eye rotates about a head-fixed horizontal axis. This supports the notion that the relatively larger contribution of the head to horizontal rotation in most situations contributes to the Fick-like range of eye-in-space (Crawford et al., 1999a).

In the second row of figure 9, the directional contributions of the head and eye have been reversed: the eye mainly rotates horizontally about a vertical axis and the head main rotates vertically (not shown) about a horizontal axis. Physically, this now resembles a Helmholtz system, where the vertical axis is embedded on a fixed horizontal axis. This results in a range of eye-in-space orientations (Fig 9-F) with an opposite twist to what we have seen so far, in other words, the opposite amount of torsion for a given gaze direction. This simulation predicts that if subjects can be induced to make gaze shifts with pure vertical head rotation, they should develop a similar range of eye-in-space orientation, unless constraints on torsion are modified in some way that has not yet been observed. This prediction could be easily tested by instructing subject to use the head vertically or horizontally in a gaze shift. In the event that people do switch to the Helmholtz constraint, this would be strong support for our model and in turn would provide an interesting experimental model for studying the influence of eye-in-space torsion on visual perception.

Thus, even if one assumes that 2D eye-head coordination and 3D eye / head constraints are implemented independently (as we have assumed here), they still interact in complex ways to influence 3D eye-in-space torsion as a function of 2D gaze direction. Since all three components of eye orientation (horizontal, vertical, and torsional) interact with 2D visual stimulus direction in a complex non-linear fashion to determine the retinal location of visual stimulation (Crawford and Guitton 1997, Henriques et al. 2000, Blohm et al. 2007), this has non-trivial implications for vision. First, it has been shown previously that the brain

accounts for 3D eye orientation in decoding patterns of visual stimulation in some behaviors (Klier and Crawford 1998; Henriques et al., 1998, Blohm and Crawford, 2007, Blohm et al., 2008), but this has not been tested in the situations simulated here. Second, it is possible that patterns of eye-head coordination are chosen to simplify or optimize patterns of retinal stimulation. Third, it is known that (contrary to the simplifying assumptions above) 3-D torsional constraints on the head are sometimes altered for different patterns of 2-D eye-head coordination (Crawford et al. 1999; Ceylan et al. 2000). This suggests the possibility that implementation of 2-D eye-head coordination and 3-D constraints might be linked in some way as to optimize vision. In short, our simulations highlight a large potential for experimental studies of the relationships between eye-head coordination and vision.

4.5 Concluding Remarks

We have proposed a kinematic model that plans accurate and coordinated eye-head gaze shifts that obey Donders' laws of the eyes and head. The following features were specifically built into the model: 1) the model transforms eye-centered retinal inputs into eye and head rotations in head and shoulder-fixed coordinate systems respectively, 2) the model applies experimentally observed behavioral constraints on the final orientations of eye (Listings law) and head (Fick strategy), and 3) variability in both eye-head contribution (including relative horizontal-vertical contributions) and influence of the VOR were implemented, without affecting the accuracy of the gaze shift or the spatial constraints named above. Our simulations show that the model was successful in realistically rendering each of these properties.

Two further novel and important properties emerged from our model simulations. First, without placing any additional dynamic constraints on the model, it predicted deviations in eye and head trajectories from the Listing and Fick between stable visual fixations that have been observed experimentally. Second, the model predicts that different patterns of eye-head coordination interact with the 3-D eye (Listing) and head (Fick) constraints to produce very different ranges of final eye-in-space orientations, with quite different consequences for vision.

Thus, our model provides both explanatory and predictive power for understanding known, and yet-to-be tested, aspects of 3-D gaze behavior. And as illustrated in Figure 1, our model provides a general framework for understanding the neural control system for the kinematics of head-free gaze control. Finally, the kinematic framework provided here provides a convenient stepping stone for further modeling studies of gaze dynamics and artificial neural network models that may further help to understand the neurophysiology of brain areas involved in gaze control.

5 General Discussion

In this thesis, we have proposed a computational framework that explains three levels of planning a gaze-shift towards cross-modal stimuli: 1) the subjects' causal inference of whether or not the cross-modal stimuli belong to a unique source, 2) the subjects' timing of making a gaze-shift towards the most reliable source of information in presence of cross-modal stimuli, 3) the kinematic coordination of eye and head movements in order to shift the line of sight. Here, in the first three sections, we discuss the results produced by our models in relation with the existing experimental observations. Then, we will describe the three models as different parts of one single executive program. Next, we will discuss the neurophysiological implications of our first two models. Then, some practical implications of the models are explained. Finally, conclusions are made and some future directions are described.

5.1 Review of Causal Inference for Multimodal Gaze-Shift Planning

We formulated the spatial causal inference problem in an executive program, i.e. we assumed that this inference is to be used to shift the focus of attention by planning a gaze-shift. This was conceptualized as choosing between three possible scenarios: 1) the signals are coming from one same object. In this case the target for gaze-shift is constructed as a weighted average of the visual and auditory estimates. 2) The signals are coming from different objects and the visual stimulus is more salient, in which case the target is chosen to be at the location of the visual stimulus. 3) The signals are coming from different objects and the auditory stimulus is more salient, so, the target is chosen to be at the location of the auditory stimulus. We realized the causal inference, within a decision making framework, through selection of one of these plans, based on the spatiotemporal similarity measure.

The proximity of the visual and auditory stimuli, in both space and time, have been experimentally shown to affect the subjects' judgement of the origin of the signals (Hairston et al., 2003, Wallace et al., 2004). The theoretical studies of causal inference have isolated these two effects leading them to reductionist models of the problem. Some have tried to model the effect of spatial disparity on the report of a common cause (Kording et al., 2007, Sato et al., 2007), ignoring the temporal dimension. Some other theoretical studies reduce the criterion for fusion to the temporal features of the events, ignore the

spatial disparity, and propose that the cross-modal events are bound together if they happen within a relative time window (Colonius and Diederich, 2010, Diederich and Colonius, 2015). However, in our model, spatiotemporal features are modeled as various dimensions of one signal.

The average percentage of the reports of a unique cause, among a number of participants and through multiple trials, changing by the spatial and temporal disparities, follow a meaningful pattern, as experimentally observed (Slutsky and Recanzone, 2001). This pattern is closely captured by the trends produced by our model, which infers the causal structure based on the spatiotemporal similarity. Unique cause is predicted for a wide range of temporal disparities if the spatial disparity is very small, as shown in figure 2A and 2C for a spatial disparity of 1.83° (ventriloquism effect). The "sameness call" changes at some point for most spatial disparities if the temporal disparity becomes greater than threshold. Similarly, the "sameness call" changes for a given temporal disparity if the spatial disparity exceeds some threshold. Thus, although we did not tinker extensively with our model parameters to exactly match the experimental results quantitatively, we conclude that the model replicates the key results and principles of the published experiment.

This general framework, which simultaneously considers spatial and temporal effects, provides a leap to understanding of other aspects of the problem. One such aspect is how different patterns of extension of presentation time of the cross-modal stimuli may change the decision. This is taken into account using the evidence-based nature of the proposed decision making circuitry, and its accumulative evolution across time. As two examples of this capability we showed that: 1) when the two stimuli are presented briefly and at the same time, they are perceived as belonging to a common source even if they are not presented at exactly the same position in space. But for the same spatial configuration, if the duration of stimulus presentation increases, the similarity measure decreases, and the decision about the uniqueness of the cause changes at some point. 2) By extending the presentation duration of one stimulus, while the other is presented only briefly, the similarity measure decreases. Therefore, the sameness decision which was for a common source for shorter durations changes to being for separate sources for longer durations.

5.2 Review of Variability of Reaction Times of Multimodal Gaze-Shifts

Previous attempts to model the variability of reaction time of saccades towards bimodal stimuli assume the temporal relationships, between the presentations of the two stimuli, as the factor governing the reaction time. Either being race models that consider two separate parallel unimodal channels (Raab, 1962, Gielen et al., 1983), or the coactivation models that consider one additive stage of processing for multimodal stimuli (Schwarz, 1989, Diederich, 1992), or the time-window-of-integration models that combine the two previous ideas, they all focus on temporal processing, ignoring the spatial effect (Frens et al., 1995). They also isolate this problem from the internal cognitive processing underlying causal inference.

Our model not only considers the effects of both spatial and temporal configurations in a dynamic network, but also relates the perceptual problem of causal inference and the executive problem of action planning in a unifying framework. Our model proposes that the patterns of variability of saccadic reaction times (RT) towards bimodal stimuli are due to high-level cognitive processing. More specifically, the decision-making process for inference of a causal structure, and the confidence on that decision, is proposed to constitute such cognitive processing. Our model explains, in a unified framework and based on cognitive assignments, the effects of a wider range of stimulus features, including spatial, temporal and reliability aspects of the stimuli, on the reaction time (Frens et al., 1995, Bell et al., 2006).

It has been observed that the reaction time of planning a gaze-shift towards cross-modal stimuli is affected by the amount of spatial distance between the visual and auditory targets, and by the temporal distance between their presentations (Frens et al., 1995, Bell et al., 2005). Our model accounts for these effects by computationally and systematically realizing the intuitive idea that the confidence on the sameness of the origin of the stimuli decreases when the spatial or temporal distance between the stimuli increases. This is accomplished in two steps: 1) introduction of the spatiotemporal similarity of the multimodal stimuli as the criterion for the causal inference, and the saliency of the multisensory plan, 2) defining confidence, driving the reaction time, as how much higher the saliency of the selected plan is relative to the other alternatives. This meant that the

plan to integrate the visual and auditory information and a gaze-shift towards their weighted average becomes less dominant relative to unimodal gaze-shift plans, when the spatial or temporal distance between the stimuli increases.

In gaze-shifts towards unimodal stimuli, it has been shown that the reaction time decreases by increasing the reliability (intensity) of the stimulus (Bell et al., 2006). Conversely, in gaze-shifts towards multimodal stimuli presented close to each other in time and space, a reduction in reaction time has been observed when the reliabilities (intensities) of the stimuli decrease (Diederich and Colonius, 2004). Our model explains both of these results based on the relative nature of the measure of confidence, introduced as the criterion for action initiation. The unisensory case occurs because of the increased confidence on a unisensory gaze-shift plan, when the stimulus intensity increases. The multisensory case happens because the dominance of the multisensory plan relative to the unisensory plans decreases when the saliencies of the unisensory plans, i.e. their stimulus reliabilities, increase. And this leads to a higher reaction time.

5.3 Review of 3D Kinematics of Head-Free Gaze-Shifts

A kinematic model was proposed for the coordinated movement of head-on-shoulder and eye-in-head in order to reorient the line of sight towards the target in the environment. The spatiotemporal eye-head coordination strategies were proposed such that the final orientations of eyes and head obey their Donders' laws, while the target is accurately foveated. The eye-centered retinal position of the target (retinal error) was used as the main input, and it was used to calculate eye and head movements relative to head- and shoulder-centered frames of reference, respectively. The free variability in the proportions of contributions of eye and head movements to gaze-shift, and the variability in the amount of vestibule-ocular eye movement, were both integrated in the model, without deviating from an accurate gaze-shift or diverging from the spatial constraint.

An accurate shift of eye orientation in space (gaze) towards the target is the main purpose of the gaze-shift machinery. This movement is trivially equal to the retinal error, for head-fixed saccades, if and only if the eyes are initially at the Listing's plane (Crawford and Guitton, 1997c, Klier and Crawford, 1998). For all other saccades, a nonlinear transformation of the retinal error is required, which depends on the initial eye and head

orientations. Failure to account for this mapping results in saccade errors that increase with the length of the retinal error. Our model does not produce such errors, and generates accurate gaze-shifts starting from any eye and head orientations. This is accomplished by systematically planned sequences of reference frame transformations between eye, head and shoulder-centered coordinates.

Spatial constraints on the orientations are of utmost importance to solve the gaze-shift problem. It has been behaviorally observed, in 3D head-fixed and head-free tasks (Glenn and Vilis, 1992, Crawford et al., 1999b), that eye, head and gaze orientations obey different forms of Donders' law. Eye orientation relative to head, at the end of gaze-shift, has zero torsional component in Euler's system (Listing's law). Head orientation relative to shoulders has zero torsion in Fick system (Fick constraint). Gaze orientation also adheres to a form of Fick-like rule. Our model directly formulates and applies the Listing's and Fick's laws on the eye and head, and shows that the Fick-like pattern shown for final gaze distributions emerge from the constraints on eye and head.

The next concern is whether or not the spatial constraints are obeyed during the temporal course of the gaze-shift. For head-fixed saccades, the Listing's law is generally believed to be obeyed (Ferman et al., 1987b, Tweed and Vilis, 1990) during the saccade, except for small torsional blips near the end of the trajectory. However, during head-free gaze-shifts, eye trajectories become much more complicated, as saccade should be coordinated with VOR, which does not obey the Listing's law (Crawford and Vilis, 1991, Klier et al., 2003). Likewise, the head deviates from the Fick range during the gaze-shift (Crawford et al., 1999b). Our model provides clear predictions for the eye, head, and gaze trajectories in 3D space, during time. Specifically, eye-in-head starts and ends the gaze-shift in the Listing's plane (LP), but deviates from LP during the gaze-shift. The starting saccade moves the eye out of the LP, while the following VOR brings the eye back to LP. The head's Fick constraint is also only applied at its initial and final orientations. The head takes the shortest path between these positions, hence violating the Fick law during the gaze-shift. Gaze trajectory, again, is not explicitly controlled, and rather is emerged from the eye and head trajectories. Gaze also deviates from its stable, quasi-Fick range during the gaze-shift.

5.4 Implications for a Complete, Multisensory Cognitive-Motor System

5.4.1 A single program of gaze-shift control, encoded in prefrontal cortex

Prefrontal cortex, at the highest level of the executive hierarchy, is involved in representing complex programs of action (Quintana and Fuster, 1999). Such programs consist of integration of multiple actions and perceptions across time, in order to achieve a goal (Petrides et al., 2012). Lesioning of prefrontal cortex causes deficits in learning to formalize action plans, by temporal integration of sensory and motor information. We would like to think of our models as parts of such a complex program of actions. While its realizing network is distributed across the cortex and subcortex, we would like to think of it as coded as a whole in prefrontal cortex. Such a unified code will activate different parts of the network at the appropriate stages of the program.

This is a program of planning a first reaction towards a possibly multimodal object in the environment. Its first element is to make the perceptual choice of whether or not the received cross-modal signals originate from a common source. This is done by constructing a spatiotemporal similarity measure and comparing it to the reliabilities of the unimodal estimates of position. The second element is to decide when to implement a gaze-shift based on the inferred cause. This is done by constructing a measure of confidence and applying a threshold on it. The third element is to move the eyes and head in a coordinated fashion spatiotemporally. This is done automatically in brainstem gaze centers.

A very important point is that an adult human develops a large number of such programs. During a real-life situation, an individual chooses one program among many. This choice is made based on the context of the situation, the type of environment, and the alertness or emotional state of the individual, among others. The program represented by our model may be used when you know multimodal information is available about the objects in the environment, and it is better to react only after you have an idea about the nature of the object. However, in similar situations, one may choose to react based on some other program of action, e.g. reacting immediately to the first stimulus appearing.

5.4.2 Attentional control and realization of working memory

Attention is an essential and inherent component of any cortical neural processing (Neisser, 1976). Every associative network of neural populations, representing and processing sensory or motor information, has as its core, the capabilities and connections to both exert attentional control and accept it. There is no evidence of a separate structure in the brain dedicated to attention as an independent function. Attention is the selective activation / deactivation of perceptual and motor networks, in a timely fashion, by some other strongly activated network, to serve the purpose of that network. When a program is executed through time, various elements of the program send attentional control signals to specific neural populations, to retain their represented signal, or process it in some special manner, through the time of the execution of the program (Fuster, 2005). What we call working memory refers to this attentional control process. The essential properties of working memory are those of a perceptual or executive memory, which is held active, in the focus of attention, as required by information processing underlying the prospective action.

We explained how we could think of the model as being coded as a unique program in prefrontal cortex. This program commands execution of different actions during time. When each element of this program is executed, a specific part of the network is chosen (attentional control) to remain active and process the corresponding information (working memory). The multisensory, short-term memory structure gets opened by the program, at first, to be updated by sensory information and to retain those signals. It then gets closed when the winning plan is sent to brainstem to drive a gaze-shift. This attentional control is applied by changing the controllable leak of this integrator. The same type of top-down attentional control is applied on the integrators that compute the spatiotemporal similarity and accumulative confidence measures. The neurons in these networks show sustained activity, similar to the signature of working memory neurons, during implementation of such programs. There are also more automatic attentional control in lower level of this program's network, an example of which is the automatic switch from "updating by the input" mode to "retaining the current content" mode in the short-term memory, based on availability of input.

5.5 Implications for Neurophysiology of Multisensory Processing

Both the causal inference and reaction time parts of the model were designed based on the known neurophysiology about multisensory integration, working memory, decision making, gaze-shift planning and action selection. The form in which this models are presented is a network of parallel processing units, whose states temporally change, similar to the structure of the brain. So this model of gaze-shift planning towards cross-modal stimuli can potentially be used to simulate spiking neural networks (Eliasmith et al., 2012) and then be compared to neurophysiological findings. Here we explain a number of instances where our models have directly realized neurophysiological findings.

5.5.1 Different levels of decision making realized in frontal cortex

We introduced three different alternative plans and three levels of decision making. The three alternative plans realize the three possible causal structures that may govern the shift of attention, and the reorientation of the line of sight. The three levels of decision making include: 1) A plan level where all the possible plans are represented. The plans that should not be considered in an instance of the task (a trial), based on the sensory context, are inhibited out of the decision making process. For example, when only visual signal is presented, the auditory and audiovisual plans are inhibited, the visual plan is the only viable one. 2) An execution level where the winning plan is disinhibited, while the other alternatives are kept under inhibition. This selective inhibition is controlled by a central decision variable that receives the bids of all viable plans and make a decision accordingly. 3) A timing level that determines when to execute the winning plan, by sending it to subcortical oculomotor machinery.

The frontal cortex includes multiple layers of neural populations with laminar organizations that are in-register representationally in each single area (Jones et al., 1977, Canteras et al., 1990, Berendse et al., 1992, Yeterian and Pandya, 1994, Levesque et al., 1996). Accordingly, the multiple alternative plans, at all the different levels of decision making, may be considered to exist in the columnar laminar structure of a single frontal area. We may choose that single frontal area to be the frontal eye fields (FEF), as it is shown to be involved in oculomotor control (Sommer and Wurtz, 2000, Wurtz et al., 2001).

A neural implementation of our model proposes that FEF represents all possible plans to guide the saccade, constructed through cognitive processing in frontal cortex, and SC represents confidence and motor maps of space at the output. FEF can possibly send any of these plans to SC for execution of a saccade. Multisensory effects in SC, then, should be interpreted as emergent property of cognitive processing in higher cortical areas reflected in plan representations in FEF. This is in line with the observation that in the lack of cortico-collicular inputs, SC neurons are incapable of integrating cross-modal signals and producing the unique response patterns (Wallace et al., 1993, Jiang et al., 2001, Alvarado et al., 2007).

5.5.2 Cortical and collicular connections of basal ganglia

A necessary feature of a neural population to qualify as a plan representation is that it can be selectively included in or suppressed from the cognitive processing. FEF neural populations are valid candidates in this respect as well, as their reciprocal connections with the basal ganglia imply (Stanton et al., 1988, Lynch et al., 1994). The implementation of a decision result can then be thought to be realized in multiple BG circuitries work in parallel (Alexander and Crutcher, 1990, Middleton and Strick, 2000). It may be proposed that, for each BG circuitry, a population of medium spiny neurons (SPN), in rostral striatum, represents the result of the decision, received from cortex. Another population of GABAergic neurons, in substantia nigra par reticulate (SNr), implements the decision result by selective disinhibition of the corresponding plan representations (Handel and Glimcher, 2000, Bayer et al., 2004). Such a model may recruit two such striato-cortical circuitries, for the implementing the first two levels of decision making. One loop may be recruited for selecting the viable plans based on the sensory context. Another loop may be recruited for selecting the winning causal structure based on the spatiotemporal similarity and the reliability of the unimodal signals.

The burst of activity in BNs of SC is dependent on a reduction of tonic inhibition in this layer. This tonic inhibition is thought to be majorly originated from the SNr (Hikosaka and Wurtz, 1985b, a, Liu and Basso, 2008). The reduction in the SNr's tonic inhibition seems to open a gate, allowing a gaze-shift plan to be communicated within SC. A neural implementation of our model may use this mechanism to guide the timing of the gaze-shift.

A SC-BG-SC loop may be proposed. The BuNs may be suggested to construct a confidence map of space. The confidence on the execution of a gaze-shift, based on the spatial position encoded, may control a GO signal that commands the BG part of the loop to stop inhibiting the BNs. Whenever this confidence ramps up above a threshold, the BNs get released out of BG's inhibition, and the spatial code is sent to brainstem to guide a gaze-shift.

5.6 Practical Implications

The kinematic part of the model proposes a strategy that may be used by the brain to coordinate oculomotor, vestibular, and head movement systems. There are a lot of patients who suffer from neck injuries, e.g. whiplash, and have problem moving their heads. Others suffer from vestibular problems like dizziness, vertigo, and lightheadedness. Our model can be used to design the internal dynamics of a training program with the goal of rehabilitation of such problems. Such training programs could, for example, be implemented in a game-like environment in virtual reality.

On one hand, as mentioned earlier in section 5.5, our model's action selection part may be neurally implemented by a model of cortico-striatal connectivity. On the other hand, Parkinson's disease has been shown to be caused by lesions or degenerations in basal ganglia neurons. A distributed neural network of BG and its projections, based on our model, may help identify neurophysiological mechanisms causing Parkinson. Also, training programs could be designed, whose internal dynamics are designed according to our model, and is used to treat Parkinson, or at least alleviate its symptoms.

A major challenge in robotics, and artificial intelligence in general, is their lack of ability to robustly interact with uncertain, changing environment. One approach to solve this problem is to develop AI algorithms whose governing equations account for temporal variabilities of their input signals and output systems. In such systems, reasoning, inference and decision making are processes whose results may change by time as signals temporally evolve. These systems account for taking different courses of action if the result of cognition changes over time. Our model is an example of such models and may be used to improve the AI systems in various industries.

5.7 Concluding Remarks

5.7.1 Conclusions

Shifting the line of sight is the major mechanism for changing the focus of attention and, consequently, updating the visual perception of the world. Such gaze-shifts may be governed by various mechanisms, e.g. the reaction to sensory perception of a stimulus in the environment, the need to update the memory of the environment, or cognitive computations based on both sensory perception and memory. One such gaze-driving mechanism was developed in this thesis for situations when sensory information from multiple modalities are presented, and the subject needs to make the most accurate gaze-shift towards the most reliable source of information. We attempted to understand how the brain plans such a gaze-shift. We proposed three parts of a unique executive program underlying this gaze-shift planning: 1) a causal inference part for identifying whether or not the signals originate from the same or different sources, 2) a timing part for determining when to implement the gaze-shift, 3) a kinematic part for coordinating eye and head movement in time and space to shift the gaze orientation.

In conclusion, the variabilities in the report of sameness, as functions of the spatiotemporal configuration of the cross-modal stimuli, result from inference of various causal structures as the source(s) of the received signals. The variabilities in the gaze-shift reaction times, as functions of the spatial temporal and reliability features of the stimuli, result from the relative confidence on the previous spatial, causal inference. The variabilities in eye-head contributions to gaze-shift and the paths they take result from the initial orientations eye and head, and also the spatial constraints applied on the stable orientations of eye and head. The final orientations and the trajectories of gaze (eye-in-space) emerge from the strategies and constraints of head-on-shoulder and eye-in-head movements.

5.7.2 Future Directions

A neural implementation of our model might shed light on the mechanisms underlying the multisensory behavior of the neurons in the superior colliculus (SC) (Stein and Stanford, 2008). This might explain the spatiotemporal principles, inverse effectiveness, and unisensory behavior of SC neurons within a framework that also explains causal inference and decision making in a multisensory task. This is achieved by associating the unit WIN

in the model to the population of saccade-related build-up neurons (BuNs), and the unit GOAL to the population of saccade-related burst neurons (BN).

The fundamental evidence that shaped neurophysiological multisensory research was about multimodal neurons in SC. They respond to cross-modal stimuli, aligned in time and space, with a significant enhancement of firing rate compared to their response to the more effective of the unisensory stimuli (Meredith and Stein, 1983, Stein et al., 1993). However, when the stimuli are presented far from each other in space or time, SC neurons show either no change or a response depression (Meredith and Stein, 1986, Frens et al., 1995). Also, the highest gain of multisensory enhancement occurs when the intensities of individual stimuli are weak. As these intensities increase the relative gain of enhancement decreases (Meredith et al., 1987, Stein and Stanford, 2008).

These findings are in direct agreement with our produced simulations of the behaviors of the units WIN and GOAL in similar tasks, and their implications about reaction times. A neural implementation of our model may explain these experimental trends by the reduction of the confidence in cortex about the notion that the stimuli come from one same source, when the spatial or temporal misalignments, or the intensities of the stimuli, increase. This can be understood by the slower build-up of activity in BuNs (representing the unit WIN), and later burst of activity in BNs (representing the unit GOAL), when the spatial or temporal distances or the intensities of the stimuli increase.

6 References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257-262.

Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13:266-271.

Alvarado JC, Stanford TR, Vaughan JW, Stein BE (2007) Cortex mediates multisensory but not unisensory integration in superior colliculus. J Neurosci 27:12775-12786.

Amlot R, Walker R, Driver J, Spence C (2003) Multimodal visual-somatosensory integration in saccade generation. Neuropsychologia 41:1-15.

Andersen RA, Snyder LH, Bradley DC, Xing J (1997) Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu Rev Neurosci 20:303-330.

Anderson JR (1983) The architecture of cognition. Cambridge, Mass.: Harvard University Press.

Anderson JR (1990) The adaptive character of thought. Hillsdale, N.J.: L. Erlbaum Associates.

Anderson JR (1995) Cognitive psychology and its implications. New York: W.H. Freeman.

Anderson JR (2007) How can the human mind occur in the physical universe? Oxford ; New York: Oxford University Press.

Angelaki DE, Dickman JD (2003) Premotor neurons encode torsional eye velocity during smooth-pursuit eye movements. J Neurosci 23:2971-2979.

Anokhin AP, Lutzenberger W, Birbaumer N (1999) Spatiotemporal organization of brain dynamics and intelligence: an EEG study in adolescents. International journal of psychophysiology : official journal of the International Organization of Psychophysiology 33:259-273.

Baddeley A (2003a) Working memory and language: an overview. Journal of communication disorders 36:189-208.

Baddeley A (2003b) Working memory: looking back and looking forward. Nat Rev Neurosci 4:829-839.

Barsalou LW (1999) Perceptual symbol systems. Behav Brain Sci 22:577-609; discussion 610-560.

Bayer HM, Handel A, Glimcher PW (2004) Eye position and memory saccade related responses in substantia nigra pars reticulata. Exp Brain Res 154:428-441.

Bechara A, Damasio H, Tranel D, Anderson SW (1998) Dissociation Of working memory from decision making within the human prefrontal cortex. J Neurosci 18:428-437.

Beiser DG, Houk JC (1998) Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. J Neurophysiol 79:3168-3188.

Bell AH, Meredith MA, Van Opstal AJ, Munoz DP (2005) Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. J Neurophysiol 93:3659-3673.

Bell AH, Meredith MA, Van Opstal AJ, Munoz DP (2006) Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Exp Brain Res 174:53-59.

Berendse HW, Galis-de Graaf Y, Groenewegen HJ (1992) Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. J Comp Neurol 316:314-347.

Berkeley G (1709) An essay towards a new theory of vision. Dublin: Printed by A. Rhames for J. Pepyat.

Bizzi E, Kalil RE, Morasso P (1972) Two modes of active eye-head coordination in monkeys. Brain Res 40:45-48.

Bizzi E, Kalil RE, Tagliasc.V (1971a) Eye-Head Coordination in Monkeys - Evidence for Centrally Patterned Organization. Science 173:452-&.

Bizzi E, Kalil RE, Tagliasco V (1971b) Eye-Head Coordination in Monkeys: Evidence for Centrally Patterned Organization. Science 173:452-454.

Blohm G, Crawford JD (2007) Computations for geometrically accurate visually guided reaching in 3-D space. J Vision 7:-.

Blohm G, Khan AZ, Ren L, Schreiber KM, Crawford JD (2008) Depth estimation from retinal disparity requires eye and head orientation signals. J Vision 8:-.

Bogacz R (2007) Optimal decision-making theories: linking neurobiology with behaviour. Trends in cognitive sciences 11:118-125.

Boring EG (1942) Sensation and perception in the history of experimental psychology. New York, London,: D. Appleton-Century Company.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci 13:87-100.

Burr D, Banks MS, Morrone MC (2009) Auditory dominance over vision in the perception of interval duration. Exp Brain Res 198:49-57.

Cannon SC, Robinson DA (1987) Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. J Neurophysiol 57:1383-1409.

Canteras NS, Shammah-Lagnado SJ, Silva BA, Ricardo JA (1990) Afferent connections of the subthalamic nucleus: a combined retrograde and anterograde horseradish peroxidase study in the rat. Brain Res 513:43-59.

Ceylan M, Henriques DY, Tweed DB, Crawford JD (2000) Task-dependent constraints in motor control: pinhole goggles make the head move like an eye. J Neurosci 20:2719-2730.

Chen L, Vroomen J (2013) Intersensory binding across space and time: a tutorial review. Attention, perception & psychophysics 75:790-811.

Chen LL, Tehovnik EJ (2007) Cortical control of eye and head movements: integration of movements and percepts. Eur J Neurosci 25:1253-1264.

Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. Nat Neurosci 11:693-702.

Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci 33:269-298.

Cohen B, Komatsuzaki A (1972) Eye movements induced by stimulation of the pontine reticular formation: evidence for integration in oculomotor pathways. Exp Neurol 36:101-117.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE (1997) Temporal dynamics of brain activation during a working memory task. Nature 386:604-608.

Cohen YE, Andersen RA (2000) Reaches to sounds encoded in an eye-centered reference frame. Neuron 27:647-652.

Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. J Cogn Neurosci 16:1000-1009.

Colonius H, Diederich A (2010) The optimal time window of visual-auditory integration: a reaction time analysis. Front Integr Neurosci 4:11.

Conklin J, Eliasmith C (2005) A controlled attractor network model of path integration in the rat. J Comput Neurosci 18:183-203.

Corneil BD, Munoz DP (1996) The influence of auditory and visual distractors on human orienting gaze shifts. J Neurosci 16:8193-8207.

Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. J Neurophysiol 88:438-454.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608-611.

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 24:87-114; discussion 114-185.

Crawford JD (1994) The oculomotor neural integrator uses a behavior-related coordinate system. J Neurosci 14:6911-6923.

Crawford JD, Cadera W, Vilis T (1991) Generation of torsional and vertical eye position signals by the interstitial nucleus of Cajal. Science 252:1551-1553.

Crawford JD, Ceylan MZ, Klier EM, Guitton D (1999a) Three-dimensional eye-head coordination during gaze saccades in the primate. Journal of Neurophysiology 81:1760-1782.

Crawford JD, Ceylan MZ, Klier EM, Guitton D (1999b) Three-dimensional eye-head coordination during gaze saccades in the primate. J Neurophysiol 81:1760-1782.

Crawford JD, Guitton D (1997a) Primate head-free saccade generator implements a desired (post-VOR) eye position command by anticipating intended head motion. J Neurophysiol 78:2811-2816.

Crawford JD, Guitton D (1997b) Primate head-free saccade generator implements a desired (post-VOR) eye position command by anticipating intended head motion. Journal of Neurophysiology 78:2811-2816.

Crawford JD, Guitton D (1997c) Visual-motor transformations required for accurate and kinematically correct saccades. J Neurophysiol 78:1447-1467.

Crawford JD, Guitton D (1997d) Visual-motor transformations required for accurate and kinematically correct saccades. Journal of Neurophysiology 78:1447-1467.

Crawford JD, Henriques DY, Medendorp WP (2011) Three-dimensional transformations for goal-directed action. Annu Rev Neurosci 34:309-331.

Crawford JD, Martinez-Trujillo JC, Klier EM (2003) Neural control of three-dimensional eye and head movements. Curr Opin Neurobiol 13:655-662.

Crawford JD, Vilis T (1991) Axes of eye rotation and Listing's law during rotations of the head. J Neurophysiol 65:407-423.

Crawford JD, Vilis T (1992a) Symmetry of oculomotor burst neuron coordinates about Listing's plane. J Neurophysiol 68:432-448.

Crawford JD, Vilis T (1992b) Symmetry of Oculomotor Burst Neuron Coordinates About Listings Plane. Journal of Neurophysiology 68:432-448.

Crutcher MD, Alexander GE (1990) Movement-related neuronal activity selectively coding either direction or muscle pattern in three motor areas of the monkey. J Neurophysiol 64:151-163.

D'Esposito M, Detre JA, Alsop DC, Shin RK, Atlas S, Grossman M (1995) The neural basis of the central executive system of working memory. Nature 378:279-281.

D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Annual review of psychology 66:115-142.

Daemi M, Crawford JD (2015) A kinematic model for 3-D head-free gaze-shifts. Frontiers in computational neuroscience 9:72.

Damasio AR (1994) Descartes' error : emotion, reason, and the human brain. New York: G.P. Putnam.

Demer JL, Oh SY, Poukens V (2000) Evidence for active control of rectus extraocular muscle pulleys. Invest Ophthalmol Vis Sci 41:1280-1290.

DeSouza JF, Keith GP, Yan X, Blohm G, Wang H, Crawford JD (2011) Intrinsic reference frames of superior colliculus visuomotor receptive fields during head-unrestrained gaze shifts. J Neurosci 31:18313-18326.

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. Exp Brain Res 91:176-180.

Diederich A (1992) Probability inequalities for testing separate activation models of divided attention. Percept Psychophys 52:714-716.

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:1388-1404.

Diederich A, Colonius H (2007) Modeling spatial effects in visual-tactile saccadic reaction time. Percept Psychophys 69:56-67.

Diederich A, Colonius H (2008a) Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model. Exp Brain Res 186:1-22.

Diederich A, Colonius H (2008b) When a high-intensity "distractor" is better then a low-intensity one: modeling the effect of an auditory or tactile nontarget stimulus on visual saccadic reaction time. Brain Res 1242:219-230.

Diederich A, Colonius H (2015) The time window of multisensory integration: relating reaction times and judgments of temporal order. Psychological review 122:232-241.

Duncan J, Burgess P, Emslie H (1995) Fluid intelligence after frontal lobe lesions. Neuropsychologia 33:261-268.

Duncan J, Emslie H, Williams P, Johnson R, Freer C (1996) Intelligence and the frontal lobe: the organization of goal-directed behavior. Cognitive psychology 30:257-303.

Edelman GM (1989) The remembered present : a biological theory of consciousness. New York: Basic Books.

Edelman GM, Mountcastle VB, Neurosciences Research Program. (1978) The mindful brain : cortical organization and the group-selective theory of higher brain function. Cambridge: MIT Press.

Eliasmith C (2005) A unified approach to building and controlling spiking attractor networks. Neural Comput 17:1276-1314.

Eliasmith C (2013) How to build a brain a neural architecture for biological cognition. pp xvii, 456 pages Oxford: Oxford University Press.

Eliasmith C, Anderson CH (2003) Neural engineering : computation, representation, and dynamics in neurobiological systems. Cambridge, Mass.: MIT Press.

Eliasmith C, Stewart TC, Choo X, Bekolay T, DeWolf T, Tang Y, Rasmussen D (2012) A large-scale model of the functioning brain. Science 338:1202-1205.

Eliasmith C, Trujillo O (2014) The use and abuse of large-scale brain models. Curr Opin Neurobiol 25:1-6.

Everling S, Fischer B (1998) The antisaccade: a review of basic research and clinical studies. Neuropsychologia 36:885-899.

Farshadmanesh F, Byrne P, Keith GP, Wang HY, Corneil BD, Crawford JD (2012a) Cross-validated models of the relationships between neck muscle electromyography and three-dimensional head kinematics during gaze behavior. Journal of Neurophysiology 107:573-590.

Farshadmanesh F, Byrne P, Wang H, Corneil BD, Crawford JD (2012b) Relationships between neck muscle electromyography and three-dimensional head kinematics during centrally induced torsional head perturbations. J Neurophysiol 108:2867-2883.

Farshadmanesh F, Klier EM, Chang P, Wang H, Crawford JD (2007) Three-dimensional eye-head coordination after injection of muscimol into the interstitial nucleus of Cajal (INC). J Neurophysiol 97:2322-2338.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Ferman L, Collewijn H, Van den Berg AV (1987a) A direct test of Listing's law--I. Human ocular torsion measured in static tertiary positions. Vision Res 27:929-938.

Ferman L, Collewijn H, Van den Berg AV (1987b) A direct test of Listing's law--II. Human ocular torsion measured under dynamic conditions. Vision Res 27:939-951.

Fetter M, Tweed D, Misslisch H, Fischer D, Koenig E (1992) Multidimensional descriptions of the optokinetic and vestibuloocular reflexes. Ann N Y Acad Sci 656:841-842.

Fodor JA, Pylyshyn ZW (1988) Connectionism and cognitive architecture: a critical analysis. Cognition 28:3-71.

Freedman EG, Sparks DL (1997) Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys. J Neurophysiol 77:2328-2348.

Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Percept Psychophys 57:802-816.

Fuchs AF, Luschei ES (1971) The activity of single trochlear nerve fibers during eye movements in the alert monkey. Exp Brain Res 13:78-89.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331-349.

Fuster JM (1997) The prefrontal cortex : anatomy, physiology, and neuropsychology of the frontal lobe. Philadelphia: Lippincott-Raven.

Fuster JM (2004) Upper processing stages of the perception-action cycle. Trends in cognitive sciences 8:143-145.

Fuster JM (2005) Cortex and mind : unifying cognition. Oxford ; New York: Oxford University Press.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652-654.

Fuster JM, Bauer RH, Jervey JP (1985) Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. Brain Res 330:299-307.

Galiana HL, Guitton D (1992) Central organization and modeling of eye-head coordination during orienting gaze shifts. Ann N Y Acad Sci 656:452-471.

Gandhi NJ, Sparks DL (2007) Dissociation of eye and head components of gaze shifts by stimulation of the omnipause neuron region. Journal of Neurophysiology 98:360-373.

Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J Neurosci 2:1527-1537.

Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. Science 233:1416-1419.

Gernert M, Hamann M, Bennay M, Loscher W, Richter A (2000) Deficit of striatal parvalbumin-reactive GABAergic interneurons and decreased basal ganglia output in a genetic rodent model of idiopathic paroxysmal dystonia. J Neurosci 20:7052-7058.

Ghasia FF, Angelaki DE (2005) Do motoneurons encode the noncommutativity of ocular rotations? Neuron 47:281-293.

Gielen SC, Schmidt RA, Van den Heuvel PJ (1983) On the nature of intersensory facilitation of reaction time. Percept Psychophys 34:161-168.

Gilbert CD (1998) Adult cortical dynamics. Physiological reviews 78:467-485.

Girard B, Berthoz A (2005) From brainstem to cortex: computational models of saccade generation circuitry. Progress in neurobiology 77:215-251.

Glenn B, Vilis T (1992) Violations of Listing's law after large eye and head gaze shifts. J Neurophysiol 68:309-318.

Godfroy M, Roumes C, Dauchy P (2003) Spatial variations of visual-auditory fusion areas. Perception 32:1233-1245.

Gold JI, Shadlen MN (2007) The neural basis of decision making. Annu Rev Neurosci 30:535-574.

Goossens HH, Van Opstal AJ (1997) Human eye-head coordination in two dimensions under different sensorimotor conditions. Exp Brain Res 114:542-560.

Groh JM, Sparks DL (1992) Two models for transforming auditory signals from head-centered to eye-centered coordinates. Biol Cybern 67:291-302.

Grossberg S (1999) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. Spatial vision 12:163-185.

Guitton D (1992) Control of eye-head coordination during orienting gaze shifts. Trends Neurosci 15:174-179.

Guitton D, Buchtel HA, Douglas RM (1985) Frontal lobe lesions in man cause difficulties in suppressing reflexive glances and in generating goal-directed saccades. Exp Brain Res 58:455-472.

Guitton D, Douglas RM, Volle M (1984) Eye-head coordination in cats. J Neurophysiol 52:1030-1050.

Guitton D, Mandl G (1980) Oblique saccades of the cat: a comparison between the durations of horizontal and vertical components. Vision Res 20:875-881.

Guitton D, Munoz DP, Galiana HL (1990) Gaze control in the cat: studies and modeling of the coupling between orienting eye and head movements in different behavioral tasks. J Neurophysiol 64:509-531.

Guitton D, Volle M (1987) Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. J Neurophysiol 58:427-459.

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. J Cogn Neurosci 15:20-29.

Handel A, Glimcher PW (2000) Contextual modulation of substantia nigra pars reticulata neurons. J Neurophysiol 83:3042-3048.

Hanes DP, Schall JD (1995) Countermanding saccades in macaque. Vis Neurosci 12:929-937.

Hanes DP, Wurtz RH (2001) Interaction of the frontal eye field and superior colliculus for saccade generation. J Neurophysiol 85:804-815.

Harrar V, Harris LR (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. Exp Brain Res 186:517-524.

Haxby JV, Petit L, Ungerleider LG, Courtney SM (2000) Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. NeuroImage 11:380-391.

Hayek FAv (1952) The sensory order : an inquiry into the foundations of theoretical psychology. Chicago: University of Chicago Press.

Healy SD, Rowe C (2014) Animal cognition in the wild. Behavioural processes 109 Pt B:101-102.

Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. Nat Rev Neurosci 9:467-479.

Helmholtz Hv, Southall JPC (1924) Helmholtz's treatise on physiological optics. Rochester, N.Y.: The Optical Society of America.

Henn V, Straumann D, Hess BJ, Hepp K, Vilis T, Reisine H (1991) Generation of vertical and torsional rapid eye movement in the rostral mesencephalon. Experimental data and clinical implications. Acta Otolaryngol Suppl 481:191-193.

Henriques DYP, Klier EM, Smith MA, Lowy D, Crawford JD (1998) Gaze-centered remapping of remembered visual space in an open-loop pointing task. J Neurosci 18:1583-1594.

Hepp K (1994) Oculomotor control: Listing's law and all that. Curr Opin Neurobiol 4:862-868.

Hepp K, Vilis T, Henn V (1988) On the generation of rapid eye movements in three dimensions. Ann N Y Acad Sci 545:140-153.

Hernandez A, Zainos A, Romo R (2000) Neuronal correlates of sensory discrimination in the somatosensory cortex. Proc Natl Acad Sci U S A 97:6191-6196.

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. Journal of experimental psychology 63:289-293.

Hikosaka O, Wurtz RH (1983a) Visual and oculomotor functions of monkey substantia nigra pars reticulata. I. Relation of visual and auditory responses to saccades. J Neurophysiol 49:1230-1253.

Hikosaka O, Wurtz RH (1983b) Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus. J Neurophysiol 49:1285-1301.

Hikosaka O, Wurtz RH (1985a) Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus. J Neurophysiol 53:266-291.

Hikosaka O, Wurtz RH (1985b) Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata. J Neurophysiol 53:292-308.

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20:583-590.

Holyoak KJ, Thagard PR (1996) Mental Leaps Analogy in Creative Thought. p 334 p. Cambridge: MIT Press.

Horak FB, Anderson ME (1984) Influence of globus pallidus on arm movements in monkeys. II. Effects of stimulation. J Neurophysiol 52:305-322.

Hummel JE, Holyoak KJ (1997) Distributed representations of structure: A theory of analogical access and mapping. Psychological review 104:427-466.

Hummel JE, Holyoak KJ (2003) A symbolic-connectionist theory of relational inference and generalization. Psychological review 110:220-264.

Iriki A, Pavlides C, Keller A, Asanuma H (1989) Long-term potentiation in the motor cortex. Science 245:1385-1387.

Jackendoff R (2002) Foundations of language : brain, meaning, grammar, evolution. Oxford ; New York: Oxford University Press.

James W (1890) The principles of psychology. New York: H. Holt.

Jay MF, Sparks DL (1987) Sensorimotor integration in the primate superior colliculus. I. Motor convergence. J Neurophysiol 57:22-34.

Jiang W, Wallace MT, Jiang H, Vaughan JW, Stein BE (2001) Two cortical areas mediate multisensory integration in superior colliculus neurons. J Neurophysiol 85:506-522.

Johnson-Laird PN (1999) Deductive reasoning. Annual review of psychology 50:109-135.

Jones EG, Coulter JD, Burton H, Porter R (1977) Cells of origin and terminal distribution of corticostriatal fibers arising in the sensory-motor cortex of monkeys. J Comp Neurol 173:53-80.

Jurgens R, Becker W, Kornhuber HH (1981) Natural and drug-induced variations of velocity and duration of human saccadic eye movements: evidence for a control of the neural pulse generator by local feedback. Biol Cybern 39:87-96.

Katus T, Grubert A, Eimer M (2015) Inter-modal attention shifts trigger the selective activation of task-relevant tactile or visual working memory representations. J Vis 15:861.

Klier EM, Crawford JD (1998) Human oculomotor system accounts for 3-D eye orientation in the visual-motor transformation for saccades. Journal of Neurophysiology 80:2274-2294.

Klier EM, Crawford JD (2003) Neural control of three-dimensional eye and head posture. Ann N Y Acad Sci 1004:122-131.

Klier EM, Meng H, Angelaki DE (2006) Three-dimensional kinematics at the level of the oculomotor plant. J Neurosci 26:2732-2737.

Klier EM, Wang H, Crawford JD (2001) The superior colliculus encodes gaze commands in retinal coordinates. Nat Neurosci 4:627-632.

Klier EM, Wang H, Crawford JD (2002) Neural mechanisms of three-dimensional eye and head movements. Ann N Y Acad Sci 956:512-514.

Klier EM, Wang H, Crawford JD (2003) Three-dimensional eye-head coordination is implemented downstream from the superior colliculus. J Neurophysiol 89:2839-2853.

Klier EM, Wang H, Crawford JD (2007) Interstitial nucleus of cajal encodes three-dimensional head orientations in Fick-like coordinates. J Neurophysiol 97:604-617.

Klink PC, Montijn JS, van Wezel RJ (2011) Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. Attention, perception & psychophysics 73:219-236.

Knudsen EI, Knudsen PF (1983) Space-mapped auditory projections from the inferior colliculus to the optic tectum in the barn owl (Tyto alba). J Comp Neurol 218:187-196.

Knudsen EI, Konishi M (1978) A neural map of auditory space in the owl. Science 200:795-797.

Koffka K (1935) Principles of Gestalt psychology. London,

New York,: K. Paul, Trench

Harcourt, Brace.

Koos T, Tepper JM (1999) Inhibitory control of neostriatal projection neurons by GABAergic interneurons. Nat Neurosci 2:467-472.

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS One 2:e943.

Kuffler SW, Nicholls JG (1976) From neuron to brain : a cellular approach to the function of the nervous system. Sunderland, Mass.: Sinauer Associates.

Kunzle H, Akert K, Wurtz RH (1976) Projection of area 8 (frontal eye field) to superior colliculus in the monkey. An autoradiographic study. Brain Res 117:487-492.

Land MF (1992) Predictable eye-head coordination during driving. Nature 359:318-320.

Levesque M, Charara A, Gagnon S, Parent A, Deschenes M (1996) Corticostriatal projections from layer V cells in rat are collaterals of long-range corticofugal axons. Brain Res 709:311-315.

Liu P, Basso MA (2008) Substantia nigra stimulation influences monkey superior colliculus neuronal activity bilaterally. J Neurophysiol 100:1098-1112.

Lochmann T, Deneve S (2011) Neural processing as causal inference. Curr Opin Neurobiol 21:774-781.

Locke J, George Fabyan Collection (Library of Congress) (1690) An essay concerning humane understanding : in four books. London: Printed by Eliz. Holt for Thomas Bassett ...

Logothetis NK, Schall JD (1989) Neuronal correlates of subjective visual perception. Science 245:761-763.

Lynch JC, Hoover JE, Strick PL (1994) Input to the primate frontal eye field from the substantia nigra, superior colliculus, and dentate nucleus demonstrated by transneuronal transport. Exp Brain Res 100:181-186.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432-1438.

Ma WJ, Rahmati M (2013) Towards a neural implementation of causal inference in cue combination. Multisens Res 26:159-176.

Maier JX, Groh JM (2009) Multisensory guidance of orienting behavior. Hear Res 258:106-112.

Markram H (2006) The blue brain project. Nat Rev Neurosci 7:153-160.

Marr D (1982) Vision : a computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman.

Martinez-Trujillo JC, Medendorp WP, Wang H, Crawford JD (2004) Frames of reference for eye-head gaze commands in primate supplementary eye fields. Neuron 44:1057-1066.

Martinez-Trujillo JC, Wang H, Crawford JD (2003) Electrical stimulation of the supplementary eye fields in the head-free macaque evokes kinematically normal gaze shifts. J Neurophysiol 89:2961-2974.

McClelland JL, Rogers TT (2003) The parallel distributed processing approach to semantic cognition. Nat Rev Neurosci 4:310-322.

McLeod P, Plunkett K, Rolls ET (1998) Introduction to connectionist modelling of cognitive processes. Oxford ; New York: Oxford University Press.

Menning H, Ackermann H, Hertrich I, Mathiak K (2005) Spatial auditory attention is modulated by tactile priming. Exp Brain Res 164:41-47.

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 7:3215-3229.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389-391.

Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res 365:350-354.

Mesulam MM (1998) From sensation to cognition. Brain 121 (Pt 6):1013-1052.

Meyer DE, Kieras DE (1997) A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. Psychological review 104:3-65.

Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev 31:236-250.

Misslisch H, Tweed D, Fetter M, Vilis T (1994) The influence of gravity on Donders' law for head movements. Vision Res 34:3017-3025.

Misslisch H, Tweed D, Hess BJ (2001) Stereopsis outweighs gravity in the control of the eyes. J Neurosci 21:RC126.

Misslisch H, Tweed D, Vilis T (1998) Neural constraints on eye motion in human eye-head saccades. J Neurophysiol 79:859-869.

Monteon JA, Avillac M, Yan X, Wang H, Crawford JD (2012) Neural mechanisms for predictive head movement strategies during sequential gaze shifts. J Neurophysiol 108:2689-2707.

Monteon JA, Constantin AG, Wang H, Martinez-Trujillo J, Crawford JD (2010) Electrical stimulation of the frontal eye fields in the head-free macaque evokes kinematically normal 3D gaze shifts. J Neurophysiol 104:3462-3475.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782-784.

Morasso P, Bizzi E, Dichgans J (1973) Adjustment of saccade characteristics during head movements. Exp Brain Res 16:492-500.

Munoz DP, Everling S (2004) Look away: the anti-saccade task and the voluntary control of eye movement. Nat Rev Neurosci 5:218-228.

Munoz DP, Guitton D (1989) Fixation and Orientation Control by the Tecto-Reticulo-Spinal System in the Cat Whose Head Is Unrestrained. Rev Neurol 145:567-579.

Munoz DP, Wurtz RH (1995a) Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. J Neurophysiol 73:2313-2333.

Munoz DP, Wurtz RH (1995b) Saccade-related activity in monkey superior colliculus. II. Spread of activity during saccades. J Neurophysiol 73:2334-2348.

Mushiake H, Inase M, Tanji J (1990) Selective coding of motor sequence in the supplementary motor area of the monkey cerebral cortex. Exp Brain Res 82:208-210.

Navarra J, Hartcher-O'Brien J, Piazza E, Spence C (2009) Adaptation to audiovisual asynchrony modulates the speeded detection of sound. Proc Natl Acad Sci U S A 106:9169-9173.

Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. Brain research Cognitive brain research 25:499-507.

Neisser U (1976) Cognition and reality : principles and implications of cognitive psychology. San Francisco: W. H. Freeman.

Newell A (1992) Precis of Unified theories of cognition. Behav Brain Sci 15:425-437.

Newell A, Simon HA (1972) Human problem solving. Englewood Cliffs, N.J.,: Prentice-Hall.

O'Reilly RC, Frank MJ, Hazy TE, Watz B (2007) PVLV: the primary value and learned value Pavlovian learning algorithm. Behavioral neuroscience 121:31-49.

Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. Nat Neurosci 14:775-782.

Pare M, Crommelinck M, Guitton D (1994) Gaze shifts evoked by stimulation of the superior colliculus in the head-free cat conform to the motor map but also depend on stimulus strength and fixation activity. Exp Brain Res 101:123-139.

Patton PE, Anastasio TJ (2003) Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. Neural Comput 15:783-810.

Petrides M, Tomaiuolo F, Yeterian EH, Pandya DN (2012) The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains. Cortex; a journal devoted to the study of the nervous system and behavior 48:46-57.

Pierrot-Deseilligny C, Rivaud S, Gaymard B, Agid Y (1991) Cortical control of memory-guided saccades in man. Exp Brain Res 83:607-617.

Pouget A, Ducom JC, Torri J, Bavelier D (2002) Multisensory spatial representations in eye-centered coordinates for reaching. Cognition 83:B1-11.

Proudlock FA, Shekhar H, Gottlob I (2003) Coordination of eye and head movements during reading. Invest Ophthalmol Vis Sci 44:2991-2998.

Quaia C, Optican LM (1998) Commutative saccadic generator is sufficient to control a 3-D ocular plant with pulleys. J Neurophysiol 79:3197-3215.

Quintana J, Fuster JM (1999) From perception to action: temporal integrative functions of prefrontal and parietal neurons. Cereb Cortex 9:213-221.

Quintana J, Yajeya J, Fuster JM (1988) Prefrontal representation of stimulus attributes during delay tasks. I. Unit activity in cross-temporal integration of sensory and sensory-motor information. Brain Res 474:211-221.

Raab DH (1962) Statistical facilitation of simple reaction times. Trans N Y Acad Sci 24:574-590.

Radeau M (1994) Auditory-visual spatial interaction and modularity. Curr Psychol Cogn 13:3-51.

Raphan T (1998) Modeling control of eye orientation in three dimensions. I. Role of muscle pulleys in determining saccadic trajectory. J Neurophysiol 79:2653-2667.

Rasmussen D, Eliasmith C (2013) Modeling brain function: current developments and future prospects. JAMA Neurol 70:1325-1329.

Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89:1009-1023.

Rieke F (1997) Spikes : exploring the neural code. Cambridge, Mass.: MIT Press.

Robertson LC (2003) Binding, spatial attention and perceptual awareness. Nat Rev Neurosci 4:93-102.

Robinson DA (1964) The Mechanics of Human Saccadic Eye Movement. The Journal of physiology 174:245-264.

Robinson DA (1973) Models of the saccadic eye movement control system. Kybernetik 14:71-83.

Robinson DA (1978) The purpose of eye movements. Invest Ophthalmol Vis Sci 17:835-837.

Rowland BA, Stanford TR, Stein BE (2007) A model of the neural mechanisms underlying multisensory integration in the superior colliculus. Perception 36:1431-1443.

Roy JE, Cullen KE (1998) A neural correlate for vestibulo-ocular reflex suppression during voluntary eye-head gaze shifts. Nat Neurosci 1:404-410.

Rubinstein L (1964) Intersensory and Intrasensory Effects in Simple Reaction Time. Perceptual and motor skills 18:159-172.

Rumelhart DE, McClelland JL, University of California San Diego. PDP Research Group. (1987) Parallel distributed processing : explorations in the microstructure of cognition. Cambridge, Mass.: MIT Press.

Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J Neurosci 29:6490-6499.

Sajad A, Sadeh M, Keith GP, Yan X, Wang H, Crawford JD (2015) Visual-Motor Transformations Within Frontal Eye Fields During Head-Unrestrained Gaze Shifts in the Monkey. Cereb Cortex 25:3932-3952.

Salinas E, Abbott LF (1994) Vector reconstruction from firing rates. J Comput Neurosci 1:89-107.

Sato KC, Tanji J (1989) Digit-muscle responses evoked from multiple intracortical foci in monkey precentral motor cortex. J Neurophysiol 62:959-970.

Sato Y, Toyoizumi T, Aihara K (2007) Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput 19:3335-3355.

Schall JD, Hanes DP (1993) Neural basis of saccade target selection in frontal eye field during visual search. Nature 366:467-469.

Schall JD, Hanes DP, Taylor TL (2000) Neural control of behavior: countermanding eye movements. Psychological research 63:299-307.

Schiller PH, Sandell JH, Maunsell JH (1987) The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. J Neurophysiol 57:1033-1049.

Schoner G, Thelen E (2006) Using dynamic field theory to rethink infant habituation. Psychological review 113:273-299.

Schreiber K, Crawford JD, Fetter M, Tweed D (2001) The motor side of depth vision. Nature 410:819-822.

Schwarz W (1989) A new model to explain the redundant-signals effect. Percept Psychophys 46:498-500.

Scudder CA, Kaneko CS, Fuchs AF (2002) The brainstem burst generator for saccadic eye movements: a modern synthesis. Exp Brain Res 142:439-462.

Shams L, Beierholm UR (2010) Causal inference in perception. Trends in cognitive sciences 14:425-432.

Singh R, Eliasmith C (2006) Higher-dimensional neurons explain the tuning and dynamics of working memory cells. J Neurosci 26:3667-3678.

Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. Neuroreport 12:7-10.

Smolensky P, Legendre Gr (2006) The harmonic mind : from neural computation to optimality-theoretic grammar. Cambridge, Mass.: MIT Press.

Sommer MA, Wurtz RH (2000) Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. J Neurophysiol 83:1979-2001.

Sparks DL (2002) The brainstem control of saccadic eye movements. Nat Rev Neurosci 3:952-964.

Sparks DL, Barton EJ, Gandhi NJ, Nelson J (2002) Studies of the role of the paramedian pontine reticular formation in the control of head-restrained and head-unrestrained gaze shifts. Ann N Y Acad Sci 956:85-98.

Sparks DL, Mays LE (1990a) Signal Transformations Required for the Generation of Saccadic Eye-Movements. Annual Review of Neuroscience 13:309-336.

Sparks DL, Mays LE (1990b) Signal transformations required for the generation of saccadic eye movements. Annu Rev Neurosci 13:309-336.

Sparks DL, Travis RP, Jr. (1971) Firing patterns of reticular formation neurons during horizontal eye movements. Brain Res 33:477-481.

Stanton GB, Goldberg ME, Bruce CJ (1988) Frontal eye field efferents in the macaque monkey: I. Subcortical pathways and topography of striatal and thalamic terminal fields. J Comp Neurol 271:473-492.

Stein BE, Meredith MA, Wallace MT (1993) The visually responsive neuron and beyond: multisensory integration in cat and monkey. Prog Brain Res 95:79-90.

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255-266.

Steinbach MJ (1987) Proprioceptive knowledge of eye position. Vision Res 27:1737-1744.

Sternberg RJ (1985) Beyond IQ : a triarchic theory of human intelligence. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Sternberg S (1969) Memory-scanning: mental processes revealed by reaction-time experiments. American scientist 57:421-457.

Stewart TC, Eliasmith C (2013) Realistic neurons can compute the operations needed by quantum probability theory and other vector symbolic architectures. Behav Brain Sci 36:307-308.

Straumann D, Haslwanter T, Hepp-Reymond MC, Hepp K (1991) Listing's law for eye, head and arm movements and their synergistic control. Exp Brain Res 86:209-215.

Straumann D, Zee DS, Solomon D, Kramer PD (1996) Validity of Listing's law during fixations, saccades, smooth pursuit eye movements, and blinks. Exp Brain Res 112:135-146.

Straumann D, Zee DS, Solomon D, Lasker AG, Roberts DC (1995) Transient torsion during and after saccades. Vision Res 35:3321-3334.

Tomlinson RD (1990) Combined eye-head gaze shifts in the primate. III. Contributions to the accuracy of gaze saccades. J Neurophysiol 64:1873-1891.

Tomlinson RD, Bahra PS (1986a) Combined eye-head gaze shifts in the primate. I. Metrics. J Neurophysiol 56:1542-1557.

Tomlinson RD, Bahra PS (1986b) Combined eye-head gaze shifts in the primate. II. Interactions between saccades and the vestibuloocular reflex. J Neurophysiol 56:1558-1570.

Tweed D (1997) Three-dimensional model of the human eye-head saccadic system. J Neurophysiol 77:654-666.

Tweed D, Haslwanter T, Fetter M (1998) Optimizing gaze control in three dimensions. Science 281:1363-1366.

Tweed D, Vilis T (1987) Implications of rotational kinematics for the oculomotor system in three dimensions. J Neurophysiol 58:832-849.

Tweed D, Vilis T (1990) Geometric relations of eye position and velocity vectors during saccades. Vision Res 30:111-127.

Uexküll Jv (1926) Theoretical biology, by J. von Uexküll. [S.l.]: [s.n.].

Ursino M, Cuppini C, Magosso E (2014) Neurocomputational approaches to modelling multisensory integration in the brain: a review. Neural Netw 60:141-165.

van der Velde F, de Kamps M (2006) Neural blackboard architectures of combinatorial structures in cognition. Behav Brain Sci 29:37-70; discussion 70-108.

van Gelder T (1998) The dynamical hypothesis in cognitive science. Behav Brain Sci 21:615-628; discussion 629-665.

van Gisbergen JA, van Opstal AJ, Schoenmakers JJ (1985) Experimental test of two models for the generation of oblique saccades. Exp Brain Res 57:321-336.

van Opstal AJ, Hepp K, Hess BJ, Straumann D, Henn V (1991) Two- rather than three-dimensional representation of saccades in monkey superior colliculus. Science 252:1313-1315.

Van Opstal J, Hepp K, Suzuki Y, Henn V (1996) Role of monkey nucleus reticularis tegmenti pontis in the stabilization of Listing's plane. J Neurosci 16:7284-7296.

Van Wanrooij MM, Bell AH, Munoz DP, Van Opstal AJ (2009) The effect of spatial-temporal audiovisual disparities on saccades in a complex scene. Exp Brain Res 198:425-437.

Van Wanrooij MM, Bremen P, John Van Opstal A (2010) Acquired prior knowledge modulates audiovisual integration. Eur J Neurosci 31:1763-1771.

Virsu V, Oksanen-Hennah H, Vedenpaa A, Jaatinen P, Lahti-Nuuttila P (2008) Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. Exp Brain Res 186:525-537.

von der Malsburg C (1995) Binding in models of perception and brain function. Curr Opin Neurobiol 5:520-526.

Vroomen J, Bertelson P, de Gelder B (2001a) Directing spatial attention towards the illusory location of a ventriloquized sound. Acta psychologica 108:21-33.

Vroomen J, Bertelson P, de Gelder B (2001b) The ventriloquist effect does not depend on the direction of automatic visual attention. Percept Psychophys 63:651-659.

Vroomen J, Keetels M (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. Journal of experimental psychology Human perception and performance 32:1063-1071.

Vroomen J, Stekelenburg JJ (2011) Perception of intersensory synchrony in audiovisual speech: not that special. Cognition 118:75-83.

Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. J Neurophysiol 69:1797-1809.

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. Exp Brain Res 158:252-258.

Wang X, Zhang M, Cohen IS, Goldberg ME (2007) The proprioceptive representation of eye position in monkey primary somatosensory cortex. Nat Neurosci 10:640-646.

Wang XJ (2008) Decision making in recurrent neuronal circuits. Neuron 60:215-234.

Weinrich M, Wise SP (1982) The premotor cortex of the monkey. J Neurosci 2:1329-1345.

Westheimer G (1959) Retinal light distribution for circular apertures in Maxwellian view. J Opt Soc Am 49:41-44.

Wurtz RH, Sommer MA, Pare M, Ferraina S (2001) Signal transformations from cerebral cortex to superior colliculus for the generation of saccades. Vision Res 41:3399-3412.

Yajeya J, Quintana J, Fuster JM (1988) Prefrontal representation of stimulus attributes during delay tasks. II. The role of behavioral significance. Brain Res 474:222-230.

Yeterian EH, Pandya DN (1994) Laminar origin of striatal and thalamic projections of the prefrontal cortex in rhesus monkeys. Exp Brain Res 99:383-398.

Young MP (1993) The Organization of Neural Systems in the Primate Cerebral-Cortex. P Roy Soc B-Biol Sci 252:13-+.

Zangemeister WH, Jones A, Stark L (1981) Dynamics of head movement trajectories: main sequence relationship. Exp Neurol 71:76-91.

Zangemeister WH, Stark L (1982) Gaze Latency - Variable Interactions of Head and Eye Latency. Exp Neurol 75:389-406.