March 5, 2022, V1

Grand Unified Theory of Mind and Brain

Part III
Holographic Visual Perception of 3D Space and Shapes

Katsushi Arisaka[1,2], Aaron Blaisdell[3,4]

University of California, Los Angeles
[1]Department of Physics and Astronomy,
[2]Department of Electrical and Computer Engineering,
[3]Department of Psychology
[4]UCLA Brain Research Institute
475 Portola Plaza, Los Angeles, CA 90095, USA
Contact: arisaka@physics.ucla.edu

Abstract

In our daily life, we visually perceive an external space and effortlessly navigate through it. Although visual stimulation reaches the 2D retinas in the egocentric frame, our brains appear to reconstruct and maintain external allocentric 3D space regardless of constantly moving eyes, head, and body. How can the 2D egocentric retinotopy be converted to the meaningful allocentric 3D space so promptly and reliably?

By applying the new concept of **Neural Holographic Tomography** (**NHT**) with **Holographic Ring Attractor Lattice** (**HAL**) developed in **Part II**, this **Part III** directly solves the above mysteries, especially the following three essential questions:

1) How are saccades compensated to establish stable vision?
2) How is 3D allocentric space visually perceived with depth?
3) How are semantic shapes recognized in a scale-invariant and rotation-invariant manner?

Fundamentally, our vision acts on the frequency-time domain by using alpha brainwaves that holographically project a predicted 3D image in front of us, very much like a 3D projection mapping onto an empty space. This procedure satisfies the basic principle of **MePMoS** in **Part I**, which is supported both causally and locally.

Table of Contents

Abbreviations

Abbreviation	Definition
AIP	Anterior Intraparietal area
AIT	Anterior Inferior Temporal cortex
CA	Cornu Ammonis
CIP	Caudal Intraparietal area
CNS	Central Nurvous System
CPG	Central Pattern Generator
CRT	Choice Reaction Time
DFT	Descrete Fourier Transformation
DG	Dentate Gyrus
ECoG	Electrocorticography
EEG	Electroencephalography
GUT	**Grand Unified Theory**
HAL	**Holographic Ring Attractor Lattice**
HC	Hippocampus
HCN	Hippocampal Network
iEEG	Invasive Electroencephalography
LEC	Lateral Entorhinal Cortex
LFP	Local Field Potential
LGN	Lateral Geniculate Nucleus
LIP	Lateral Intraparietal cortex
LTD	Long-term Depression
LTP	Long-term Potentiation
MEC	Medial Entorhinal Cortex
MEG	Magnetoencephalography
MePMoS	**Memory-Prediction-Motion-Sensing**
MIP	Medial Intraparietal cortex
MST	Medial Superior Temporal area
MT	Middle-Temporal Area
NHT	**Neural Holographic Tomography**
PFC	Prefrontal Cortex
PIT	Posterior Inferior Temporal cortex
PN	Pulvinar Nuclei
RF	Receptive Field
RSC	Retrosplenial Cortex
RT	Reaction Time
SAC	Somatosensory Association Cortex (Brodmann Areas 5 and 7)
SC	Superior Colliculus
SRT	Simple Reation Time
STDP	Spike Timing-Dependent Plasticity
SWR	Sharp Wave and Ripples
TRN	Thalamic Reticular Nucleus
VTC	Ventral Temporal Cortex

1 Human 3D Vision – Review of Parts I & II

1.1 Introduction – Unsolved Mysteries of Human Vision

Our daily lives rely heavily on vision. Nearly half of our brain is dedicated to handling visual signals, but the true origin of vivid visual sensations of external 3D space is still a substantial mystery. How can we perceive external stable allocentric 3D space? An external object is visually perceived where it is physically located in 3D. We then observe 3D-shaped objects. This visual reconstruction in 3D is so natural that we do not usually realize its non-triviality, but why is the image projected out there, rather than inside our brain? In this **Part III** of the **Grand Unified Theory of Mind and Brain**, we shall focus on possible accurate solutions, following the new theoretical models developed on top of the previous two parts:

Part I: Space-Time approaches to Dynamic Connectomes of *C. elegans* and Human Brains
Part II: Neural Holographic Tomography (NHT) and Holographic Ring Attractor Lattice (HAL)

Let us first review the known mysteries of vision. One cannot deny that the external world we perceive is 3D and allocentric, filled with numerous meaningful landmarks with semantic shapes and descriptive characteristics like color. Thus, through evolution, our visual signal processing must have been optimized to recognize 3D allocentric space faithfully and with its associated embedded landmarks. Such visual perception seems so effortless that one might think of it as natural and insignificant. However, further consideration tells us its non-trivial mysterious ability.

First, we seem to reconstruct external 3D space even though the visual image on the retina is fundamentally 2D. Second, we constantly move our eyes unconsciously by saccades, but our visual perception is stable and unaltered. Even after extensive vision research over the past few decades, these elemental mysteries have not yet been solved.

The other puzzle is our extraordinary ability to recognize 3D shapes of complex landmarks, such as human faces and cars, and extract their semantic meanings in a fraction of a second. Suppose we memorize only one landmark with one apparent size and a fixed orientation at its object-centric frame. In order to compare the memory of the landmark directly with a new incoming image, the visual pathways must perform the seven degrees of frame translation described below swiftly, either on the memory or on the incoming image, until they overlap and agree with each other:

1) 3D parallel translation between the egocentric sensation and the object-centric memory
2) Scaling up/down to match the size of the memory and the sensory image.
3) 3D rotation of [Roll, Pitch, Yaw] to align the memory and the sensory image orientation.

Traditional bottom-up approaches from sensor inputs cannot handle these 7D frame translations at all. What is necessary is a top-down approach, starting from **Me**mory and **P**rediction. The top-down concept was initially proposed by Singer and Gray (1995). We already explored it extensively in **Part I** (Arisaka 2022a) and developed a new idea of **MePMoS (Me**mory-**P**rediction-**Mo**tion-**S**ensing). This top-down concept satisfies the fundamental physics law of causality and locality; thus, it is consistent with Hebbian plasticity. Furthermore, in **Part II** (Arisaka 2022b)**,** we invented a new model of **NHT** (Neural Holographic Tomography), resulting in a specific structure of the engram, named **HAL** (Holographic Ring Attractor Lattice) that can be considered as an *engram*.

This **Part III** is structured as follows. In the remaining **Section 1**, we quickly review **Part I** and **Part II**, emphasizing the evolutionary origin and unified principle of 3D navigation and 3D vision. Then **Section 2** will present an overview of the human-specific visual pathways: the ventral "where" pathway and the dorsal "what" pathways. **Section 3** will examine how the ventral pathway extracts and recognizes semantic shapes like human faces. **Section 4** is devoted to the holographic principle of

depth perception, resulting in conscious 3D visual perception of external space. **Section 5** will focus on the dorsal pathway that reconstructs and maintains 3D visual perception under the body-centric space, utilizing overt attention and covert attention, followed by **Section 6** that describes the conversion from the body-centric Linear-polar to the truly allocentric Cartesian coordinate system. Finally, **Section 7** will present the latest evidence discovered by our various Reaction Time (RT) experiments. [*The remaining **Section 1** reviews **Part I – II**. Readers who have read them and become familiar with the contents are advised to skip them and directly go to **Section 2**.*]

1.2 Review of MePMoS and Dynamic Space-Time Connectomes

In **Part I**, we proposed that the fundamental principle of physics, causality and locality, must be imposed strictly at every synaptic connection. This is such a strong constraint that textbooks endorsing the traditional view of neuroscience must be rewritten from scratch. Even though the brain's primary purpose is to perceive external space, this requirement tells us that neurons cannot perceive space by simply spiking based on bottom-up stimulation. Let's take an example of a retinotopic image in our visual cortex. The so-called Receptive Field (RF) is "invisible" as long as these neurons in the RF flash at random timing. Instead, to perceive any image (2D space and shape), all the neurons in the RF must flash coherently by assigning the well-ordered time sequence.

The critical consequence is that the brain is fundamentally a top-down organ, acting according to the signal flow of **Me**mory → **P**rediction → **Mo**tion → **S**ensing, named **MePMoS**. Only if spatial information is well predicted from memory in advance by the time sequence, and at the same time if sensory stimulation also expresses spatial information in temporal sequence, these two temporary patterns can be directly compared through time coincidence to generate "visible" perception, because it is the only way to satisfy causality and locality. In our **MePMoS** model, various low-frequency brainwaves (theta/alpha/beta, $f < 30$ Hz) play an essential role in assigning such tight coincidence by phase coherence. The most fundamental brainwave is the theta brainwave (~5 Hz) for decision-making per cycle, while visual perception is coordinated by the alpha brainwave (~10 Hz).

Figure 1-A (taken from **Part I: Section 5.2-3**) illustrates two theta cycles of the **MePMoS** model. Let's assume that, at t = 0 ms, we must take an initial guess (without sensory input) and decide to move in a specific direction (shown as **Me**). After that, at t~100 ms, a new sensory signal comes in and undergoes a comparison to the brain's prediction, completing **MePS** at t ~200 ms. During the second theta cycle (t = 200 – 400 ms), we have a much better prediction since we are taking the previous sensory input into account. Finally, after the two theta cycles (~400 ms), **MePMoS** is completed. The end of the first theta cycle (~ 200 ms) corresponds to unconscious attention, whereas the end of the second one (~400 ms) generates conscious awareness. Following the above flowchart, we completed the whole space-time diagram of the human brain, shown in **Figure 1-B**, with emphasis on visual perception of 3D space and shape. This **Part III** will untangle the detailed functions behind this complex diagram, step by step. [*More detailed treatment can be found in* **Part I**.]

1.3 Review of NHT (Neural Holographic Tomography) for 3D Vision

In **Part II**, we explored the general principle of space-to-time conversion for both 3D navigation and vision. We started with an insect navigation system which utilizes a ring attractor. Their ring attractor exhibits a remarkable function of 1D → 1D' linear frame translation (in the polar coordinate system) to maintain the allocentric frame. By taking it, we proposed a new concept of **Neural Holographic Tomography** (**NHT**) to conduct 3D (space) → 1D (space) + 2D (time). Once the spatial dimension is reduced from 3D to 1D, any linear translation in 3D can be achieved by 1D → 1D' space translation with holographic 2D time shifts in the frequency-time domain.

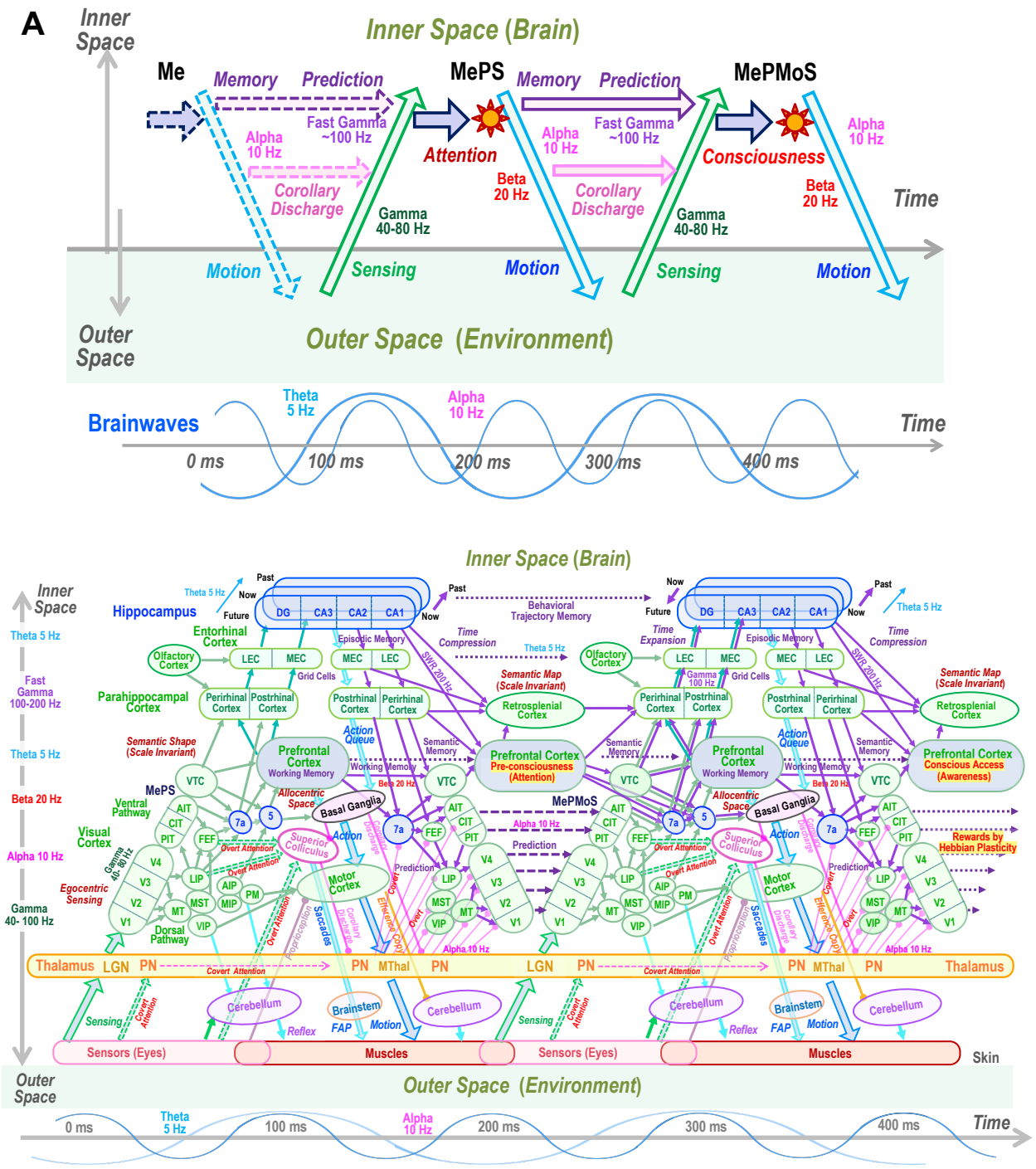

Figure 1. **(A)** An overall space-time flowchart of the human brain, based on **MePMoS**. Two cycles of **MePMoS** explain the origin of attention (at ~200 ms) and conscious awareness (at ~400 ms). 200 ms is a period of the theta brainwave (*f* ~5 Hz), segmented into two cycles of the alpha brainwave (*f* ~10 Hz). **(B)** A complete dynamic connectome of the human brain expanded from (A). Emphasis is given to visual signal processing and sensory-motor integration for perception and memory of 3D space and shape. [*Both figures are taken from* **Part I: Section 5.2-3**.]

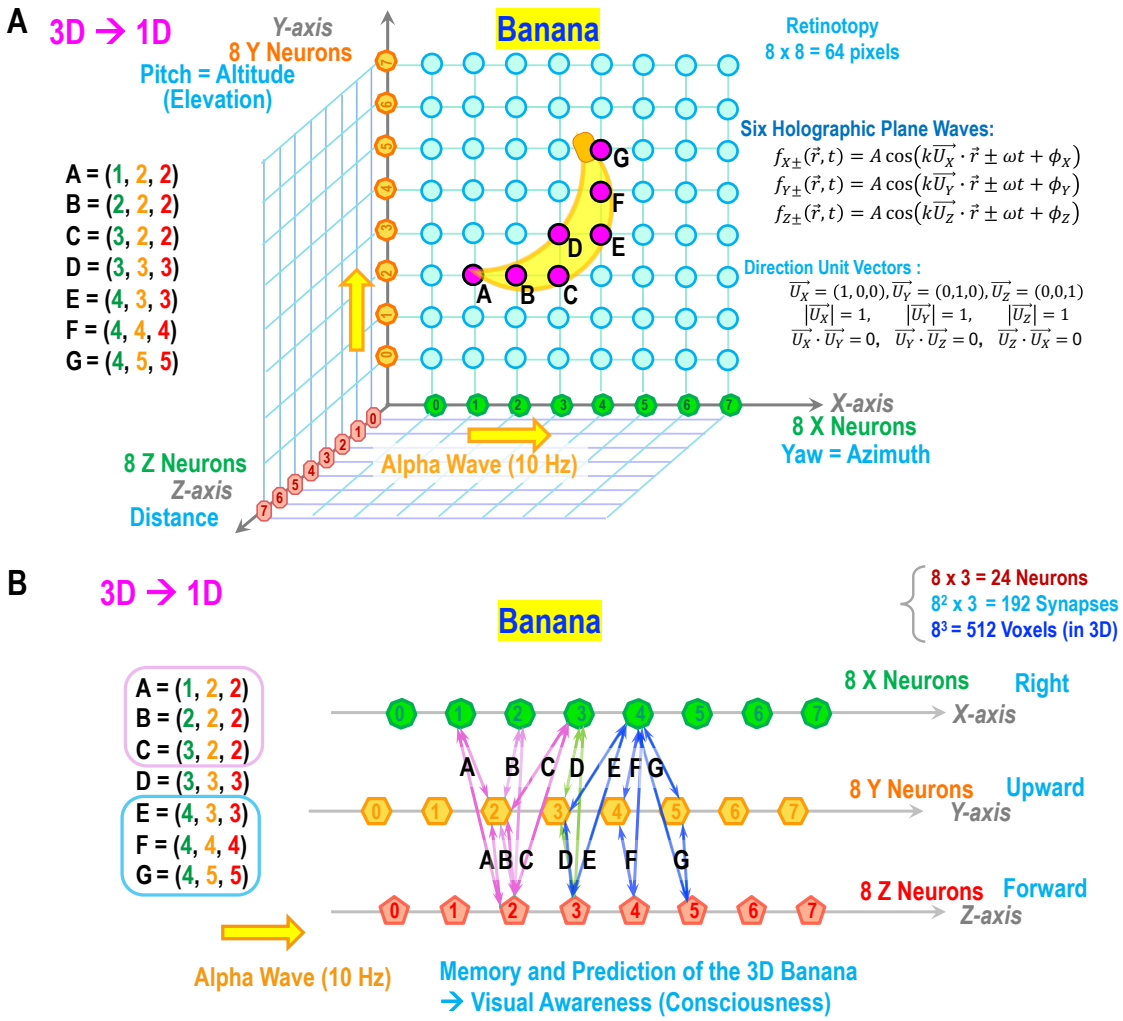

Figure 2. Principle of **Neural Holographic Tomography (NHT)** for 3D vision. **(A)** 3D toy model for 3D vision is recording a 3D banana shape, recorded by 7 points from A to G. **(B)** demonstrates a compact arrangement of the memory unit established by the 7 (points) x 3 (combinations) x 2 (dual directions) = 42 synaptic connections. [*Both figures are taken from* **Part II: Section 3.2**.]

Figure 2 (taken from **Part II: Section 3.2**) shows an example of visual perception of a banana shape in 3D (on an 8 x 8 = 64 retinotopic matrix). By sending three alpha brainwaves along the three axes, three linear neurons (8 each) can encode the 3D banana shape by 1D (space) + 2D (phase of alpha). The 3D shape can be stored by the matrix of synapses among the three lines of neurons as shown in **Figure 2-B**, resulting in long-term static memory of the 3D banana shape. Such a compact of form can be considered as an *engram*.

The essential advantage of this synaptic memory is that it is independent of the brainwave that generates it. Therefore, it can be retrieved by any frequency with any phase. In particular, the flexibility of the phase shifts allows for performing the linear frame translation in 3D, including the one between the allocentric frame (of perception and memory) and the egocentric frame (of sensing). This is the fundamental principle behind our stable visual perception of external allocentric 3D space.

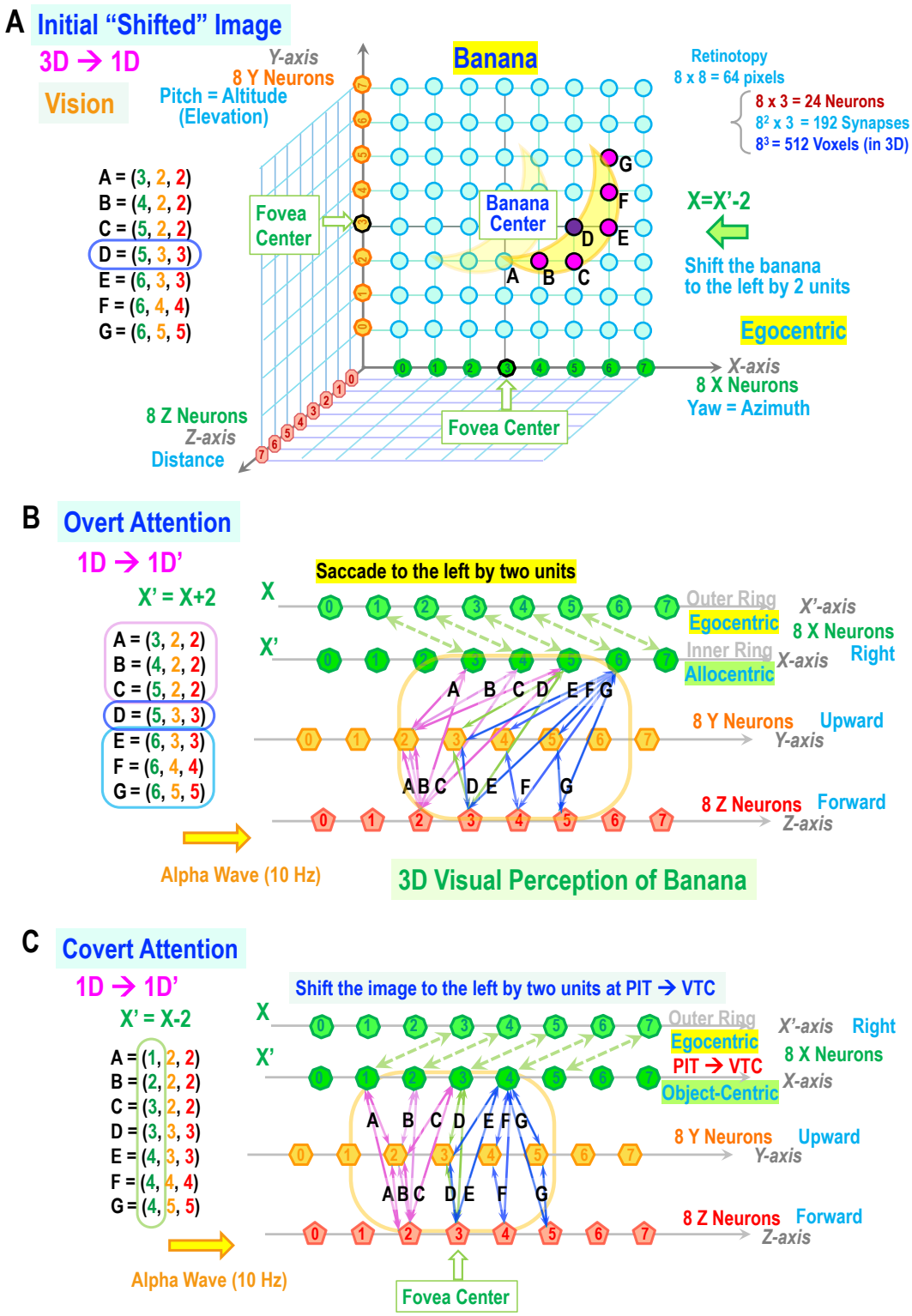

Figure 3. The holographic origin of overt and covert attention. **(A)** shows a retinotopic image of the banana that is slightly shifted horizontally from **Figure 2** to the right by two pixels. **(B)** illustrates how overt attention recognizes the banana shape while maintaining the allocentric frame unchanged. **(C)** explains how covert attention works to extract the semantic banana shape without a saccade. [*Three figures are taken from* **Part II: Section 3.3**.]

1.4 Review of Overt and Covert Attention

Why is our visual perception so stable despite constant saccadic eye movements? The holographic concept of **NHT** can immediately solve this problem in one shot. Let's assume we observe a banana slightly off-centered to the right, as shown in **Figure 3-A**. We can take two possible actions: overt attention or covert attention.

In overt attention, we unconsciously reallocated the fovea center to the banana center, D = (3, 3, 3), but the perceived banana stays at the exact allocentric location as before. This is because the newly centered banana image is shifted back to the allocentric frame by the horizontal shift of X' = X+2, as shown in **Figure 3-B** (like the insect's ring attractor for navigation.). In covert attention, eyes are stationary, but the off-centered banana shape is transferred by X' = X-2 through the dorsal pathway via the alpha phase shift. As shown in **Figure 3-C**, we can now directly compare it with the memorized shape in **Figure 2**. [*More detailed explanation was given in **Part II: Section 3.3***.] We will revisit it later in **Section 5** after the dorsal pathway is presented in detail.

1.5 HAL (Holographic Ring Attractor Lattice) for 3D Vision

The above example of the 3D banana shape (**Figures 2** and **3)** is a simplified toy model to demonstrate the essence of **NHT**, but it is overly simplistic. So, in **Part II**, we introduced a more advanced specific model, **Holographic Ring Attractor Lattice (HAL**), to define a realistic memory unit for 3D navigation and 3D vision**. Figures 4** and **5** are the proposed **3D Vision HAL** based on the 3D polar coordinate system of [Yaw, Pitch, Roll, Distance].

HAL is a hypothetical compressed 2D lattice array consisting of 16 (neurons) x 4 (dimensions) x 2 (lines) = 128 neurons, as shown in **Figure 5-C**. The **3D vision HAL** for our conscious visual perception is the outcome of the dorsal pathway. It is represented by the Linear-polar coordinate systems given in **Figure 4-A**. Three axes of [Yaw, Pitch, Distance] by the Linear-polar system asymptotically behave like the Cartesian system at long distance (>> a few meters away) near the central view (< 10 degrees): [Yaw, Pitch, Distance] ~ [X, Y, Z]. In contrast, the ventral pathway utilizes the 2D Log-polar coordinate system formed by the two axes of [Roll, Log(Eccentricity)]. In **Figures 4** and **5,** we propose the combined **HAL** representation by the four parameters [Yaw, Pitch, Roll, Distance]. Here, the fourth axes could be assigned to either Distance, Log(Distance), or Log(Eccentricity). After the two visual pathways are introduced, we will examine these options for the fourth axis, later in **Sections 3-5**.

Please note that in **Figure 4**, the Caesarian coordinate of [X, Y] is also assigned in the place of [Yaw, Pitch]. This is because our 3D visual perception in daily life seems to generate a supplementary sensation (= visual awareness) of the allocentric frame by the true 3D Cartesian coordinate system of [X, Y, Z]. We especially experience this during the navigation toward a landmark, which gives the critical input to the Hippocampal network. So, at the end of the visual pathway around the region 7a, the linear polar coordinate of the body-centric 3D vision, [Yaw, Pitch, Roll, Distance], is naturally converted to the allocentric 3D Cartesian coordinate, [X. Y, Z], which enters the Parahippocampal cortex. We will examine this transition in **Section 6** and **Part IV**.

Through **Section 1**, we have reviewed the essential new concepts of **MePMoS**, **NHT**, and **HAL** from **Part I** and **Part II**. By combining and applying these, we have already developed the basic model of human 3D vision based on holographic tomography. To our best knowledge, this **HAL**-based 3D vision is the first model that satisfies causality and locality at every single synapse level. It also defines the exact structure of the *engram*. In the following sections, we will address the significant mysteries of human vision below.

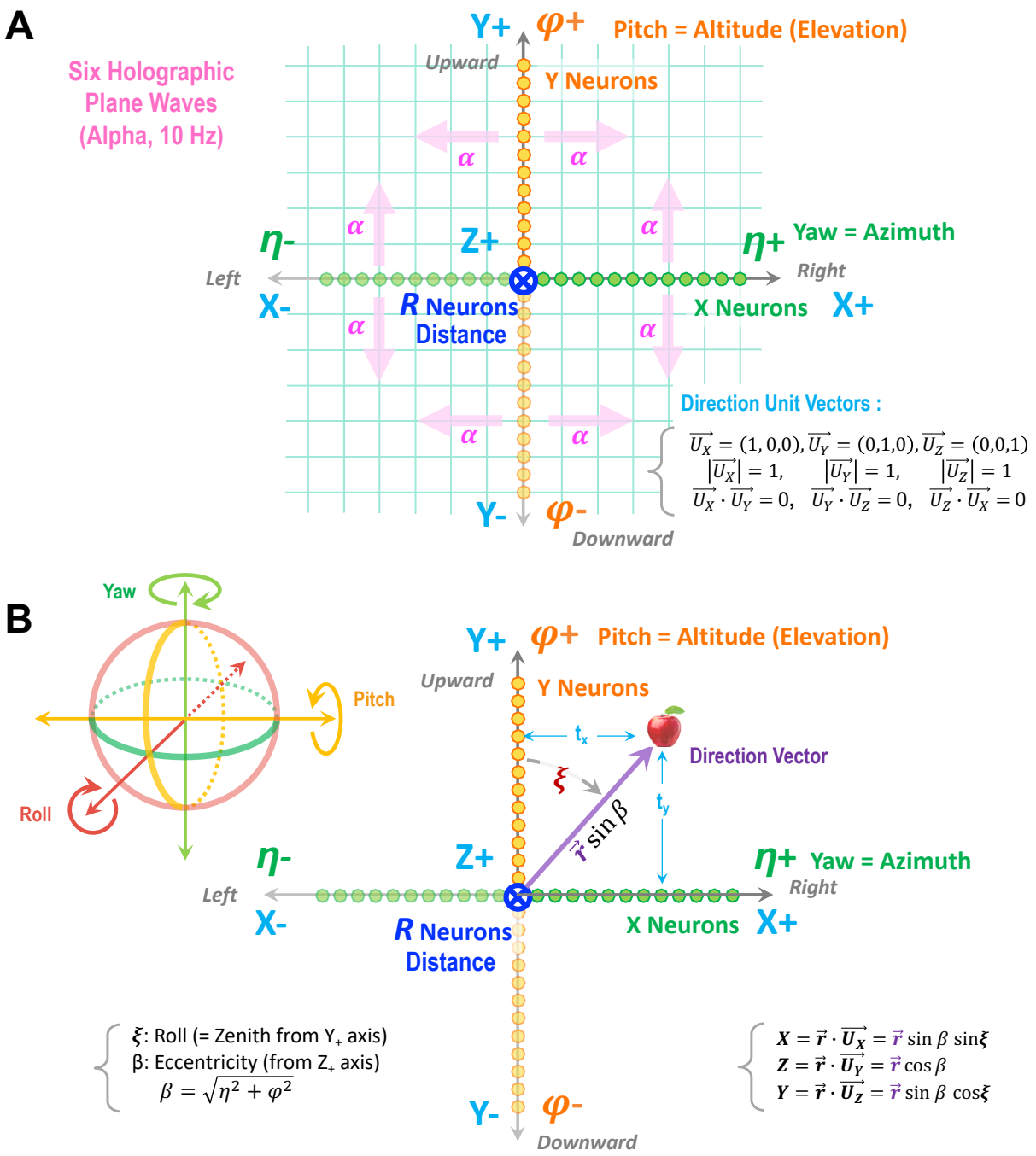

Figure 4. 3D Vision HAL by the alpha brainwave. **(A)** A total of six alpha waves are traveling in the positive and negative directions of [Yaw, Pitch, Distance]. **(B)** illustrates how the location of a landmark (= an apple) is registered by this **3D Vision HAL**. To be exact, our 3D vision has two distinct representations: **3D Linear-Polar HAL** by [Yaw, Roll, Distance (*R*)] and the true **Cartesian HAL** by [X, Y, Z]. The transition from the Linear-polar to Cartesian will be explained in **Section 6**. [*These figures were not included in* **Part II: Section 5.3**, *but they are consistent.*]

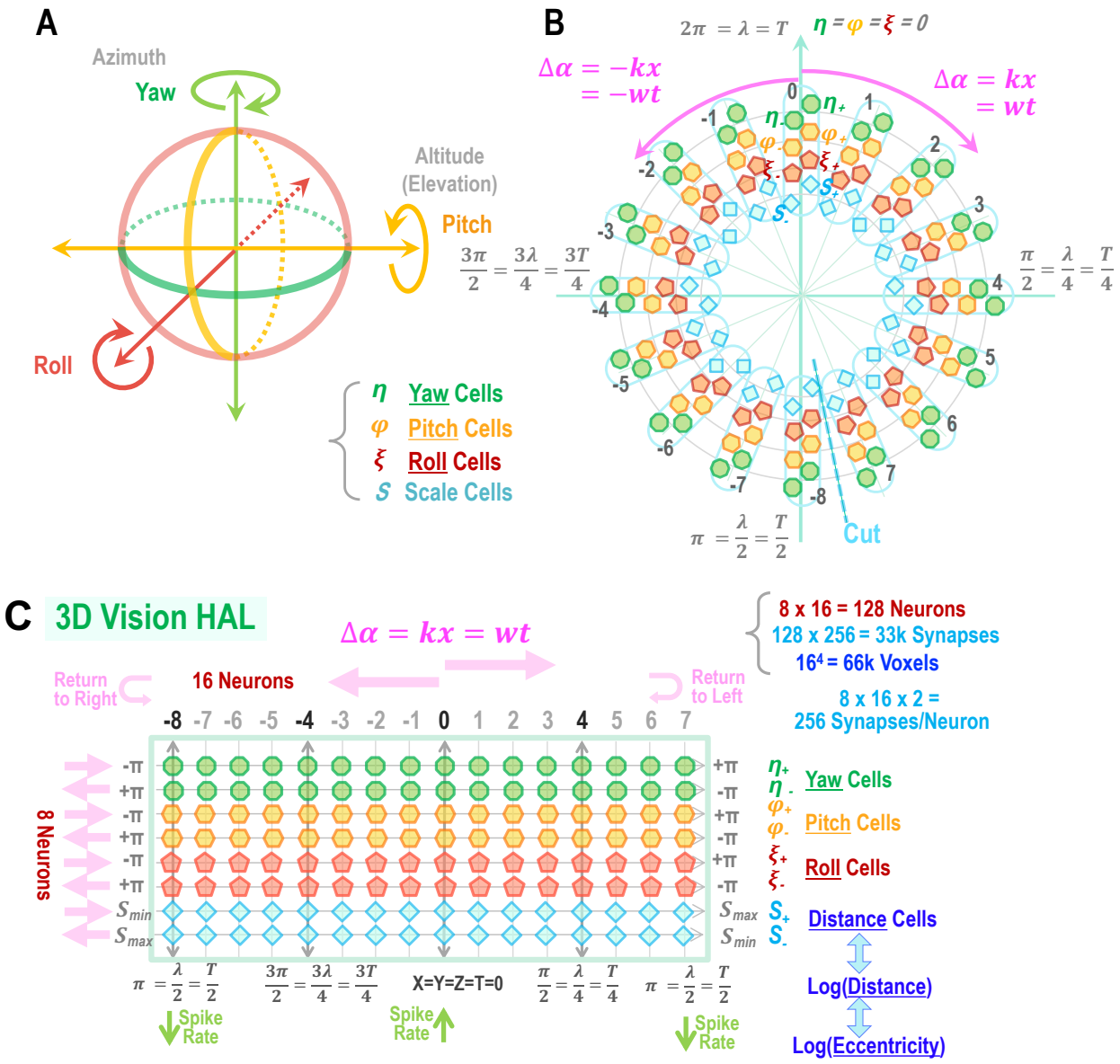

Figure 5. 3D Polar Vision HAL by the alpha brainwaves. **(A)** defines the 3D rotation in [Yaw, Pitch, Roll]. **(B)** shows the 3D Ring Attractor. Three rotations of [Yaw, Pitch, Roll] are scanned by a total of the six beta brainwaves, running at both clockwise and anti-clockwise directions. The fourth corresponds to distance. **(C)** is the **Holographic Ring Attractor Lattice (HAL)** for 3D vision. With 8 horizontal lines with 16 neuron segments each, a **HAL** contains 8 x 16 = 128 neurons. Each neuron is mutually connected with all other neurons in dual ways. [*These figures are from* **Part II: Section 5.3**.]

1) How we recognize the different sizes and orientations of shapes (like human faces): **Section 3**
2) How we perceive depth**: Section 4**
3) How we maintain stable vision regardless of saccades: **Section 5**
4) How we obtain a visual perception of allocentric 3D space**: Section 6**

But before we move on to these Sections, in the following **Section 2**, we will review the current understanding of the ventral and dorsal visual pathways in detail.

2 Dual Visual Pathways for Perceiving 3D Space and Shape

2.1 Ventral and Dorsal Visual Pathways in the Human Visual System

We shall finally begin the new contents of **Part III: Holographic Visual Perception of 3D Space and Shape**. As a starting point, let us develop the dynamic space-time diagram of the human visual system. The human brain is morphologically a complex 3D organ, where sophisticated neural networks are running as a function of time. Therefore, ultimately, we must investigate the 4D space-time diagram in (x*, y, z, t*)**.** Unfortunately, since a sheet of paper (or a computer display) is a flat 2D surface, a traditional practical approach is to take a static 3D (*x, y, z*) structure of the brain and to project it onto a 2D (*x, y*) plane, ignoring the depth (*z*) and time (*t*) dimensions, as shown in **Figure 6-A**.

In contrast, our unique approach is the space-time diagram of the brain in (*x', t*) already given in **Figure 1-B**, where the 3D volumetric structure in (x*, y, z*) is compressed to the 1D vertical axis (= *x'*) while keeping the 1D time axis (t) horizontally. Such a 2D expression in space-time is the convention of particle physics, named the "Feynman" diagram. For clearer inspection, the parts of visual pathways in **Figure 1-B** are extracted and enlarged in **Figure 6-B,** only for the first theta cycle (0 – 200 ms) of visual signal processing.

As we discussed repeatedly, adding the time axis to the neural network diagram in Figures **1-B** and **6-B** is the heart of our new model, **MePMoS** and **NHT**. By integrating **Figures 6-A** and **B** conceptually, we should be able to imagine and unveil the actual 4D dynamic brain in (*x, y, z, t*), which is the central theme of this article. To make sure, in **Figure 6-B**, the primary visual cortex, V1/V2/V3/V4, appears three times at *t* = 0 - 50, 100 - 150, 200 - 250 ms because it acts differently on the three stages of image processing in the order of (1) bottom-up sensing (= attention), (2) top-down prediction from memory, and (3) the second stage of bottom-up sensing for confirmation (= conscious visual awareness). Please note that the **MePMoS** starts from the second stage in this diagram at *t* = 100 ms.

Let us explore **Figure 6-B** in detail. Following a traditional approach, we decided to start this figure with an unexpected visual stimulus at *t* = 0. From *t* = 0, the bottom-up sensory signals ascend towards the primary visual pathway in the following order of eyes (Retina) → Thalamus LGN → Primary visual cortex: V1 → V2 → V3/V4. At this point, the signals branch out to the ventral and dorsal pathways; the ventral pathway goes like V4 → PIT → CIT → AIT→ VTC (Fusiform), whereas the dorsal pathway goes through MT → MST/LIP/VIP → FEF → 7a. The detailed connectome of these visual pathways was first published by Felleman and Essen (1991) and then further advanced by several later studies (Gilbert & Li, 2013; DiCarlo, Zoccolan, & Rust, 2012; Kruger et al., 2013).

The above bottom-up process takes place in 0 – 100 ms. So far, we are following the conventional approach. However, as we extensively discussed, the retinotopic bottom-up signals by themselves are "invisible" because they are still "space-like" and do not communicate with each other in the time domain. Thus, they cannot generate "conscious" visual perception. According to **MePMoS**, the sensory **S**ignal must be compared with **Me**mory → **P**rediction → **Mo**tion directly by time coincidence. The top-down process of **MePMo** is shown in the 100 – 200 ms time range in **Figure 6-B**. This part is a mirror image of the 0 – 100 ms range by flipping it in time like 100 → 0 ms, but at the same time, to maintain the causality and locality, all the arrows of signals must still go to the future to the right, resulting in the shown top-down (left-right) arrows.

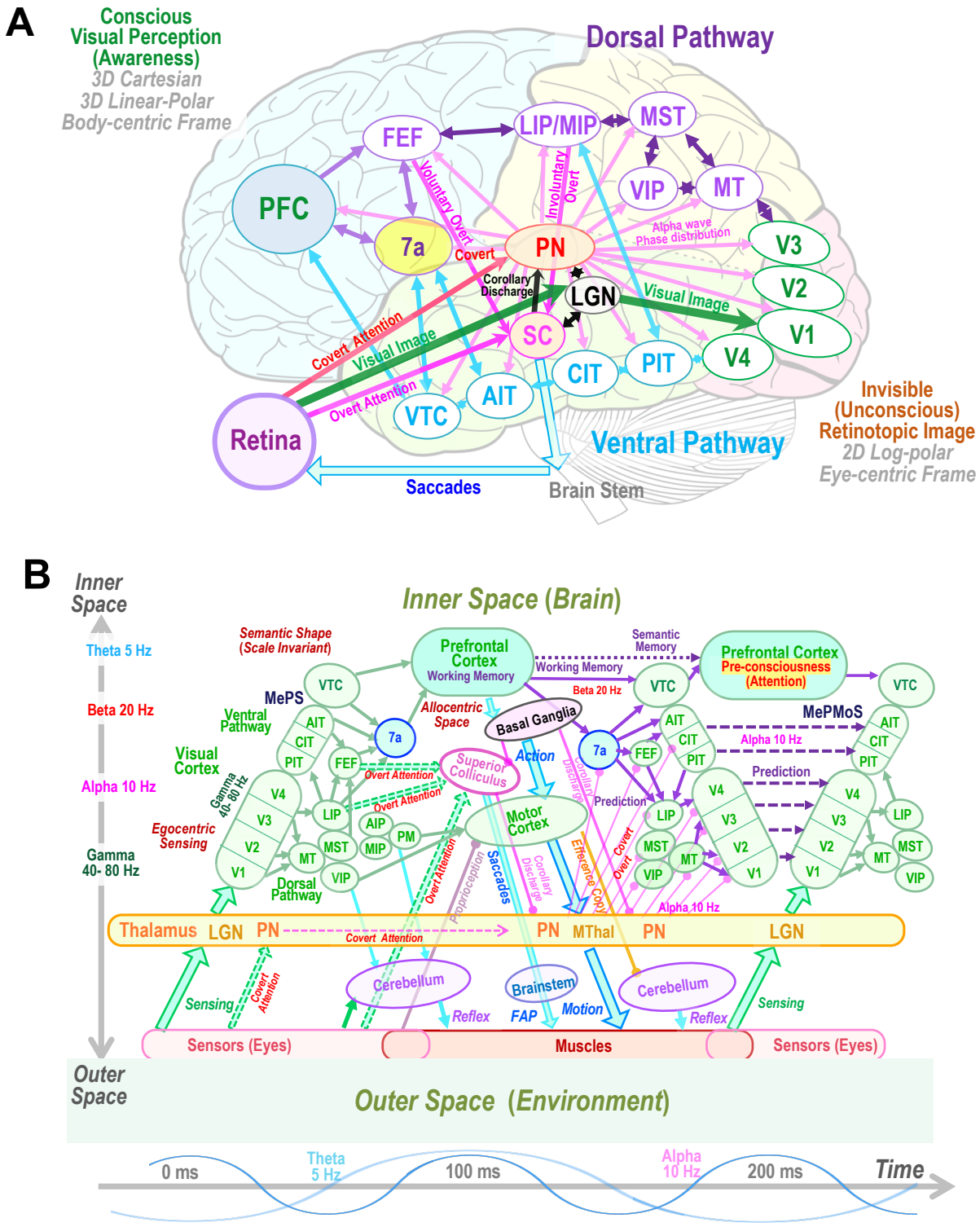

Figure 6. Illustrations of the visual pathways in space only and space-time. **(A)** shows the geometric static diagram of the visual system in (x, y) on the cross-section of the human brain. **(B)** A space-time diagram of human visual pathways, extracted **from Figure 1-B**. From an unexpected visual stimulus at t = 0, bottom-up signals go towards the upper-right through the visual pathways. Then, at t = 100 m in return, top-down signals branch out to the visual pathways for prediction. [*These figures are taken from* **Part I: Section 5.3**.]

Among these arrows, the dark purple lines are **Me**mory → **P**rediction starting from the Prefrontal Cortex → 7a → to the entire dorsal and ventral pathways all the way down to V4/V3/V2/V1 (in the reverse order as 0 → 100 ms). The thick cyan arrows are **Mo**tion from Basal Ganglia → Motor Cortex → Motor Thalamus → Muscles. (The motion could be a simple one or complex sequence as shown here.) The combination of the above forms the **MePMo**.

2.2 PN Network in Frequency-Time Domain for NHT

How can the diagram ensure time coincidence between the **MePMo** and **S**ensing at the second theta cycle starting at *t* ~200 ms? The essence is shown by the additional numerous pink lines in **Figures 6-A** and **6-B** (with dots at the end) in *t* = 150 – 200 mm. From the Basal Ganglia (BG) → Superior Colliculus (SC) → Pulvinar Nucleus (PN), the corollary discharge is injected as a faithful efferent copy of **Mo**tion (Bridge, Leopold, & Bourne 2016; Soares et al., 2017). In our **NHT** model, this corollary discharge is distributed everywhere in the visual pathways as the new phase assignments of the alpha brainwaves.

The essence is that the PN acts as the central hub, like a master clock of a CPU, to distribute brainwaves to the entire cortex with appropriate phases. This is doe location by location to match the coordinate systems there, as shown by the numerous pink lines radially spread from the PN to everywhere, which occupies 100 – 200 ms in **Figure 6-B**. This is an inhibitory synaptic network associated with gap junctions. Thanks to these proper alpha phase shifts, when the second cycle of sensing signals arrives in the 200 – 250 ms range, **MePMoS** is realized by the precise coincidences everywhere in the visual cortex down to V1 and up to the Prefrontal cortex, which is the essential principle of the **MePMoS**.

Besides the above visual signal processing, from the top-down and bottom-up, two more parallel signal paths co-exist for covert and overt attention as bypass networks (Spering and Carrasco 2015). As shown in **Figure 6-A**, a saccadic eye movement by overt attention can be generated involuntarily by the bypass bottom-up processing via Retina → Superior Colliculus (SC) → Brainstem (Engbert 2006; Martinez-Conde et al. 2013). Once saccade proceeds from SC, a proper corollary discharge is created and propagated via SC → PN. Conversely, in the case of covert attention, signals from the retina directly go to PN bypassing the SC. Therefore, no saccadic motion is initiated but new attention is initiated at PN to recognize a shape in peripheral vision (see **Sections 5.3** and **5.4**.)

The accurate distribution of brainwaves with specific phases is the heart of executing the **MePMoS**. The basic concept of phase assignment from the PN is derived from (Coulon and Landisman 2017; Llinas 2014). Llinas showed that the higher membrane potential of neurons in the PN increases the frequency of brainwaves and vice versa (Llinas 2014). It implies that if the potential goes down below the threshold for a moment, the brainwave would be shut off momentarily during that period. Such a quick turn-off of the brainwave effectively increases the phase. According to Coulon (Coulon and Landisman 2017), the cortical-wide distribution of the exact frequencies and phases is regulated by gap junctions.

Please note that the phase distribution by the PN is fundamentally a top-down process. But to be exact, it may be more suitable to call it a "**feed-out**" process. And if we decide to name it so, the corollary discharge (or efference copy) from SC → PN could be named "**feed-in.**" Then, we would better categorize the signal processing by the four processes: bottom-up, top-down, feed-in, and feed-out.

Due to the real-time feed-out of brainwave distributions with specific phases from PN, the actual top-down signals of **Me**mory → **P**rediction can properly go down to both the dorsal and ventral pathways. The dorsal pathway ensures this by converting the memorized and perceived allocentric (body-centered) 3D frame back to the 2D egocentric retinotopy through the top-down pathway: PFC

→ 7a → FEF → LIP/MIP → MST/VIP → MT → V3 → V2 → V1. Likewise, the ventral pathway propagates the predicted semantic shapes down to the ventral pathway in the reverse order: VTC (Fusiform) → AIT → CIT → PIT → V4 → (V3/V2) → V1.

Through these complete processes in both dorsal and ventral pathways, and due to the precise alpha phase assignments, the top-down signals and the bottom-up signals can handshake with each other under the proper coordinate systems at each specific location. It is this handshake of **MePMoS** that establishes the mutual communication and coordination to execute the 7D frame translation between the perception at PFC and the sensing at the primary visual cortex V1. It is not too surprising that the combination of the top-down (from PFC) and the feed-out (from PN) contributes to ~90% of the actual connections in the visual systems. On the other hand, the bottom-up (from V1) only contributes to ~10% of the connections (Gilbert and Li 2013).

Figure 6-A also illustrates the similarity and differences between overt and covert attention. As already mentioned, overt attention goes from Retina → SC → Brainstem for a saccade, and its corollary discharge goes from SC → PN → everywhere. On the other hand, covert attention bypasses SC and directly goes from Retina → PN. In this case, the dorsal pathway does not have to perform the remapping (as eyes are not moving), but the attended eccentric image must be transformed to the object-centered coordinate system along the ventral pathway before it reaches the VTC. This image transformation is achieved at 7a in our model. Detailed remapping mechanisms by covert and overt attention will be given later in **Section 5.3** and **Section 5.4** respectively.

The ventral pathway shows well-defined retinotopy in V1 → V2 → V3, but then it abruptly disappears at the PIT. This is where the **NHT** must take place and the 2D retinotopy is deducted to 1D (space) + 1D (time), utilizing the traveling alpha wave. After PIT, the object-centered holographic image is formed for further scaling and "Roll" rotation to recognize its shape.

2.3 Dual Coordinate Systems of Visual Pathways and Perception

Fundamentally, the primary purpose of the above visual signal processing is to reconstruct the accurate expression of the external allocentric world based on egocentric sensory inputs. Therefore, the consolidation between the allocentric and egocentric frames is the key to a successful operation. The distribution of various coordinate systems in the human visual cortex is illustrated in **Figure 7**.

Roughly speaking, the frontal side of our brain in PFC consciously perceives the 3D allocentric frame by the 3D Linear-polar frame of [Yaw, Pitch, Distance], which is eventually converted to the true Cartesian frame of [X, Y, Z] and injected into the Hippocampal network for 3D navigation. To be exact, in the cortical region of 7a, we speculate that the holographic coordinate system is represented as a 4D expression of [Yaw, Pitch, Roll, Distance]. This is because the outcome of the dorsal pathway must be [Yaw, Pitch, Distance], while the ventral pathway should be [Roll, Log(Eccentricity)]. Later in **Section 4**, we will show that the last axis of **HAL** can be expressed by either Distance, Log(Distance), or Log(Eccentricity), which is critical for depth perception. The right side of **Figure 7** displays the bottom-up sensory signals expressed by the egocentric 2D retinotopy of [Roll, Log(Eccentricity)] which starts from LGN → V1 → V2 → V3/V4.

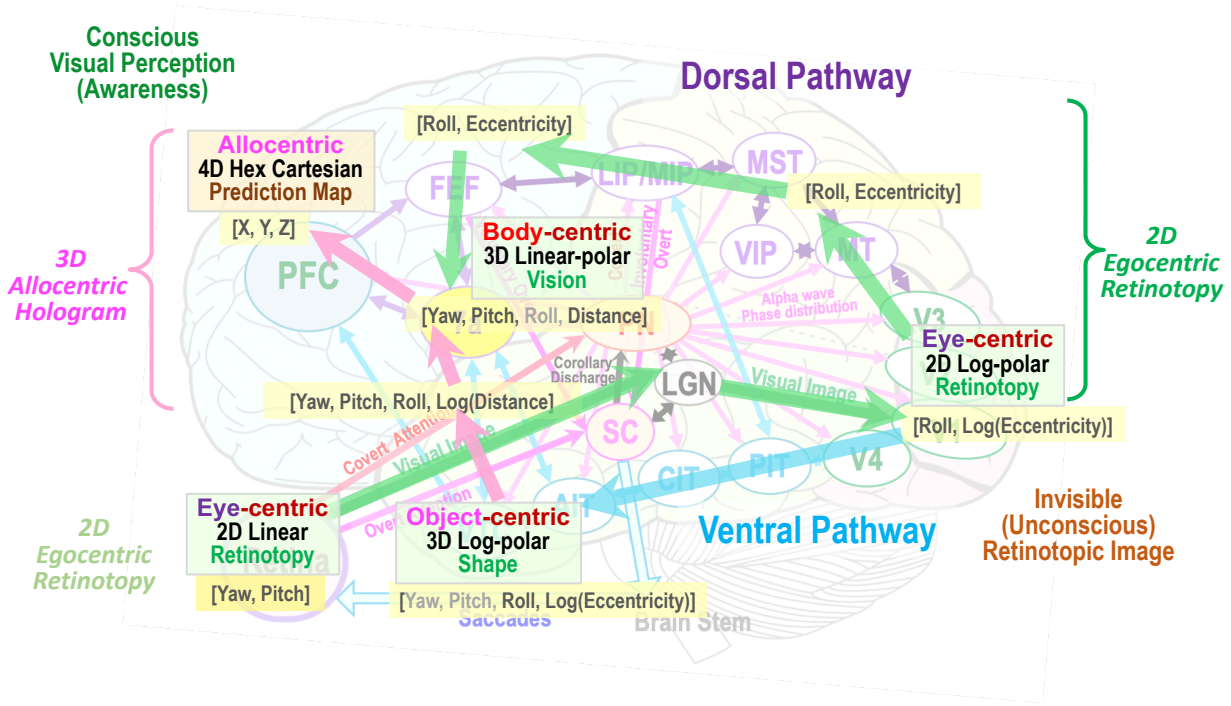

Figure 7. Representation of all types of the 2D and 3D coordinate systems in our visual system, superimposed on top of **Figure 6-A**. Three types of egocentric frames – Ege-centric, Body-centric, and Object-centric frames – are distributed.

Function	Location in Brain	2D / 3D	Coodinate	Brain-wave	Center	Direc-tion	Axes of HAL	Scale
Vision								
2D Image	Eyes	2D	2D Linear	Retino-topy	Eye		[Yaw, Pitch]	N/A
2D Receptive Field	Primary (LGN→V1/V4)		Log-Polar				[Roll, Log(Ecc.)]	Log(Ecc.)
3D Space Construction	Dorsal (MT→FEF)	2D →3D	Linear-Polar				[Roll, Ecc., Log(Dist.)]	Log(Dist.)
3D Shape Recognition	Ventral (PIT→VTC)		Log-Polar	Alpha	Object		[Roll, Log(Ecc.), Log(Dist.)]	
Conscious 3D Vision	7a → Parahippo-campal Cortex	3D	Linear-Polar		Body		[Yaw, Pitch, Roll, Dist.]	Distance
Conscious 3D Space			Cartesian		Alloc.		[X, Y, Z]	N/A
Navigation								
Head Direction	(Everywhere)	3D	Linear-Polar	Beta	Head	Alloc.	[Yaw, Pitch, Roll, Speed]	Speed
Path Integration - In	Parahippocampal Cortex				Object		[Yaw, Pitch, Roll, Dist.]	Distance
			Cartesian	Theta	Alloc.		[X, Y, Z]	N/A
Map-based Navigation	EC-Hippocampus						[X', Y', Z', T] Hex tilted	N/A
Path Integration - Out	Retrosplenial Cortex		Log-Polar	Beta	Object		[Yaw, Pitch, Roll, Log(Dist.)]	Log(Dist.)

Table 1. A complete list of the existing coordinate systems in the human brain for both vision and navigation in 3D. Since the visual system goes to the navigation systems, these two share the common coordinate systems for the polar and Cartesian. Please note that the coordinate systems must specify the center of the axes and the direction of the primary axis separately. [*This table is from* **Part II: Section 4.1**.]

We have already reviewed the existing coordinate systems of the brain in **Part II: Section 4, Table 1**, which is duplicated here as **Table 1** again**.** This table displays various coordinate systems from left to right, which are categorized in the columns below:

3D Navigation (Map-based vs. Path integration) vs. 3D Vision

1) Cartesian vs. Polar coordinate systems
2) Linear coordinate by (X, Y, Z) vs. Polar coordinate system by [Yaw, Pitch, Roll]
    a. Within the Polar coordinate: Linear-polar vs. Log-polar
3) Brain locations responsible for expressing the holographic coordinates.
4) Type of brainwave: Theta (~5 Hz), Alpha (~10 Hz), and Beta (~20 Hz)
5) 3D map center and 3D direction: Allocentric vs. Egocentric
    a. Within Egocentric: Object-centric, Eye-centric, Head-centric, and Body-centric
6) 1D scale factor: distance, Log(Distance), Log(Eccentricity), and Speed.

Among these, conscious 3D visual perception is based on the polar coordinate system of [Yaw, Pitch, Roll], which is nearly identical to 3D navigation by path integration. But to be exact, vision is body-centric whereas navigation is allocentric. In unconscious visual signal processing, both Linear-scale and Log-scale representations appear to co-exist, which is a unique feature that is only observed in the human brain. [*We will address this issue in* ***Section 6***.]

From the bottom of **Table 1**, the visual signals go through several stages following the bottom-up process; After the ventral pathway, the 3D semantic shape is constructed by the log-polar coordinate system at VTC (Fusiform). On the other hand, along the dorsal pathway from MT to FEF, 3D space is constructed by the linear polar coordinate system, finally at 7a.

To be exact, visual stimulation goes up following this order while changing coordinate systems:

1) Eye-centric 2D Linear by [Yaw, Pitch] on the retina
2) Eye-centric 2D Log-polar by [Roll, Log(Eccentricity)] on the primary visual cortex (V1→ V4)
3) Object-centric 2D Log-polar [Roll, Log(Eccentricity)] along the ventral pathway (PIT→VTC)
4) Eye-centric 2D Linear by [Roll, Eccentricity] along the dorsal pathway (MT → FEF)
5) Body-centric 3D Linear-polar [Yaw, Pitch, Distance] (7a, after saccade correction)

Why are there so many different coordinate systems in our vision? Between the Linear-polar vs. Log-polar, Linear-polar is nearly space-invariant in [Yaw, Pitch], so it is critical for overt/covert attention for the proper prompt linear frame translation. In contrast, Log-polar is scale and roll invariant, thus it is essential for 2D shape recognition by scaling and rotation. Among the Polar and Cartesian, the Polar coordinate requires only one scaling factor. It is also decoupled from the direction vector. On the other hand, the Cartesian requires three scaling factors to be applied to the three axes. Therefore, it is not suitable for the operation of scaling.

2.4 Complete Space-time Diagram of Visual Pathways

Lastly, **Figure 8** shows the detailed functional space-time diagram of the visual pathways. **Figure 8-A** follows the traditional bottom-up diagram from the visual pathways up to PFC and the hippocampal network (HCN). In contrast, **Figure 8-B** is a mirror image of **Figure 8-A,** showing the top-down signal pathways, starting from PFC and HCN, down to the primary visual cortex.

Figure 8 includes the complete connectome that has been reported so far: (Gilbert and Li 2013; Spering and Carrasco 2015; Kruger et al. 2013; Bridge et al. 2016; Soares et al. 2017; Martinez-Conde et al. 2013; Engbert 2006; Miller and Buschman 2013; Freedman and Ibos 2018). It is completely consistent with **6-B** as it should be; **Figure 8-A** corresponds to t = 0 – 100 ms in **Figure 6-B** for the bottom-up **S**ensory signal processing, whereas **Figure 8-B** corresponds to t = 100 – 200 ms in **Figure 6-B** for the top-down prediction, which is **MePMo**.

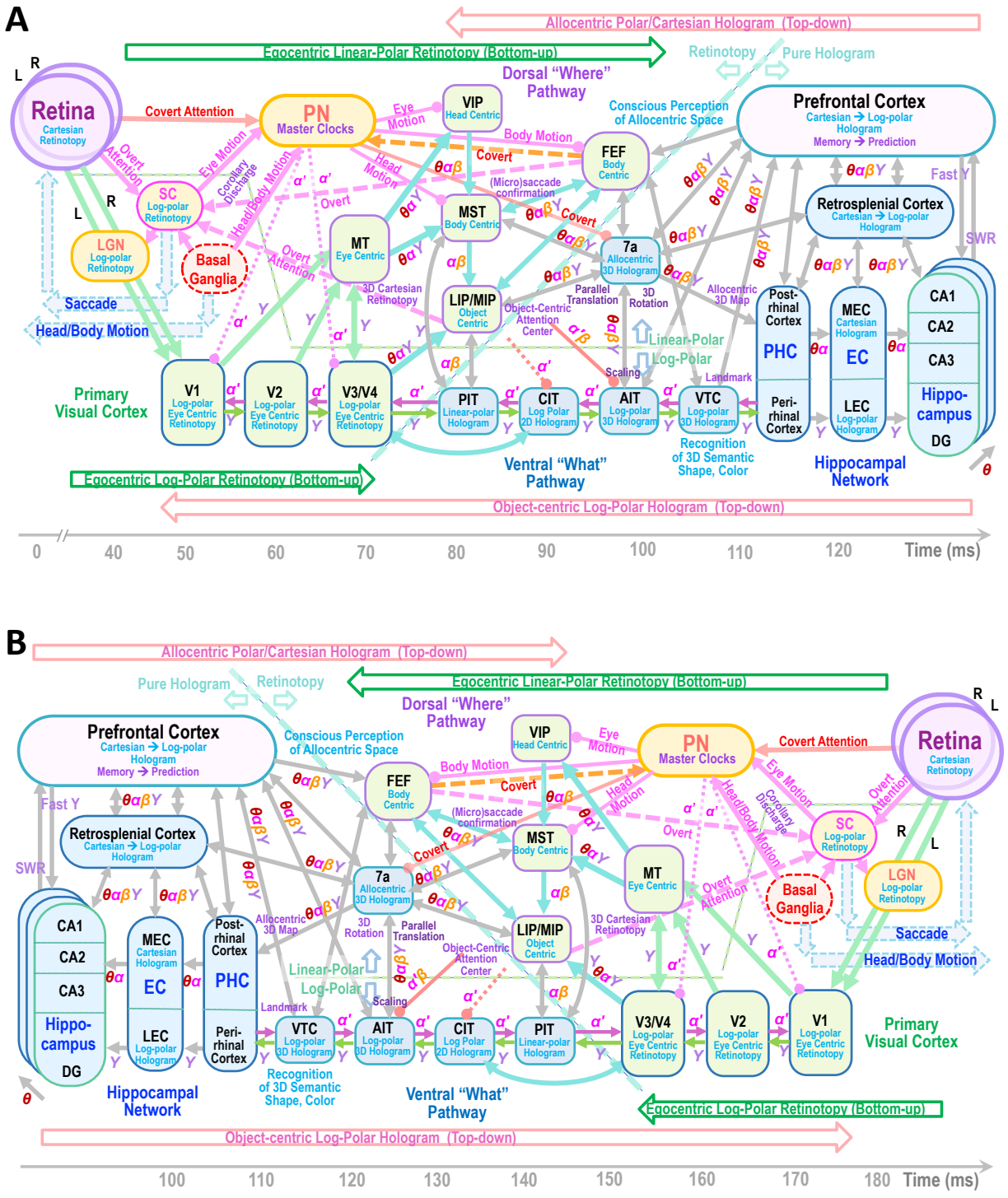

Figure 8. Detailed functional space-time diagram of the visual pathways. **(A)** shows the traditional bottom-up diagram from the visual pathways up to PFC and the hippocampal network (HCN). **(B)** is a mirror image of (A), showing the top-down signal pathways, starting from PFC and HCN down to the primary visual cortex. These two diagrams act together through time management by the alpha wave phase assignment via PN.

As discussed already, these two diagrams these two diagrams act and handshake together through time management by the alpha wave phase assignment via PN, feed-in and feed-out, realizing the **MePMoS** in order of **Figure 8-B → 8-A**.

Figure 8-A is divided into left and right by the 45^{o}-tilted dashed line (cyan) near the center, which separates the 2D retinotopy (to the left) from the pure holographic expression (to the right). The retinotopy is "invisible" because of its "space-like" nature, whereas the hologram can be consciously visible because it is "time-like". The translation among various coordinate systems is only possible under holographic expression, which is the essence of **NHT** and **HAL** models.

In the following sections, we will utilize **Figures 6**, **7**, and **8** to investigate the accurate functions of each part of the ventral and dorsal pathways, working together via the PN networks.

In summary, human visual pathways are arguably the most complex brain-wide networks to conduct 3D frame conversion from the egocentric to allocentric frames. At the same time, semantic information must be extracted promptly and assigned to key landmarks. Nevertheless, with the new concepts of MePMoS, NHT, and HAL, we are beginning to unveil the hidden fundamental principles of their operations. In the following sections, we will finally attack the remaining profound mysteries of our vision below:

1) How we recognize the different sizes and orientations of shapes (like human faces): **Section 3**
2) How we perceive 3D visual space with depth**: Section 4**
3) How we maintain stable vision regardless of saccades: **Section 5**
4) How we generate a visual perception of allocentric 3D apace**: Section 6**

3 Recognition of 2D Shape by the Ventral Pathway

3.1 Ventral "What" Pathway and 2D Shape HAL

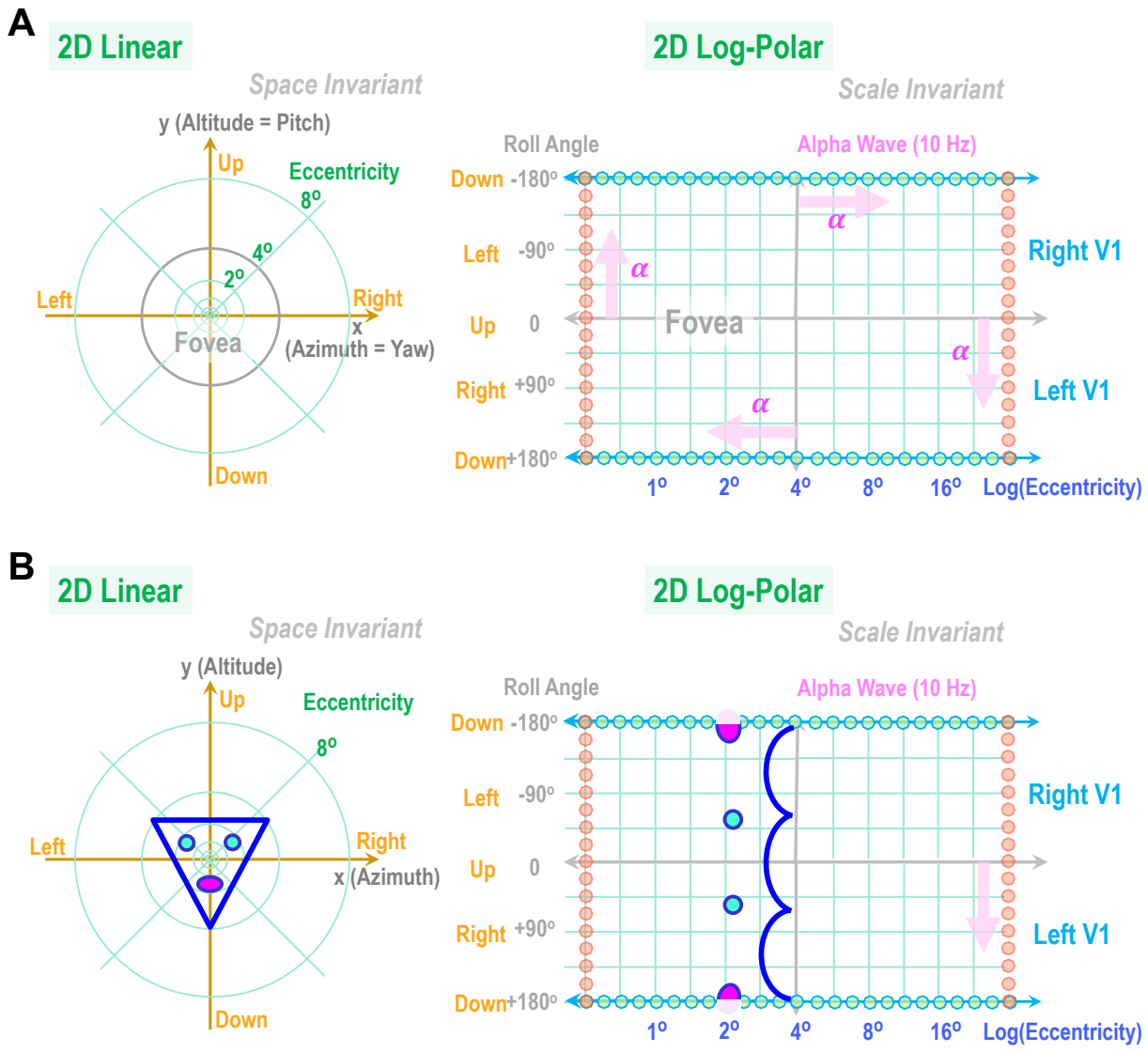

Figure 9. Frame Translation from retinotopy under the Cartesian coordinate system to the Log-polar Coordinate System, which is observed in human V1-V4. **(A)** Expression of both coordinates; The left side is the 2D Linear coordinate by [Yaw, Pitch] = [Azimuth, Altitude] on the retina, whereas the right side is the Log-polar coordinate of V1. The fovea is the central range within ~5º of eccentricity, which is mapped to the left half of the Log-polar V1. **(B)** An example of frame conversion of a hsimplified human face from the Linear to Log-polar. Please note that the nose at the fovea center is extremely enlarged on the Log-polar.

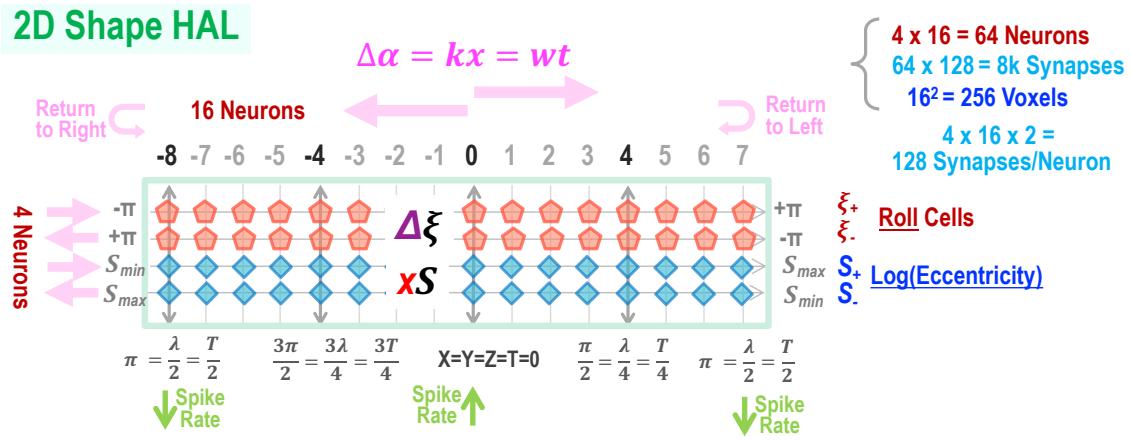

Figure 10. 2D Shape HAL of the Ventral Visual Pathway. Since the standard HAL is initially designed to represent 4D space-time, only 1D + 1D lines for [Roll, Log(Eccentricity)] are required for the **2D shape HAL**. The central vertical line (X = Y = Z= T = 0) corresponds to the upward direction for the Roll Cells, and Eccentricity = 5° for Scale Cells of Log(Eccentricity), thus the left side corresponds to the fovea. The Alpha brain brainwaves start from the center with zero phases, and go out to the right and left line by line.

In this **Section 3**, we shall address the first mystery of vision; How can we recognize various shapes, such as human faces and cars, regardless of their apparent different sizes and rotating angles? This is a non-trivial issue that has been overlooked in the past. The traditional approach by the Receptive Field (RF) through the bottom-up process cannot address this question at all. Firstly, the retinotopic RF is "invisible" because it is "space-like". Secondly, even if the RF is converted to the frequency-time domain by alpha brainwaves, if the size is different from the memorized one, we cannot directly compare them by time coincidence because they do not overlap with each other. Therefore, as we discussed repeatedly, we need a radically new top-down approach by integrating the concepts of **MePMoS**, **NHT**, and **HAL** altogether.

Let us first investigate the bottom-up signal processing along the ventral pathway, known as the "what" pathway. It goes like Retina → LGN → V1 → (V2/V3) → V4 → PIT → CIT → AIT → VTC (Fusiform), as shown in **Figure 8-A**. Initially, the retinal image is under the 2D Linear coordinate system of [Yaw, Pitch] = [Azimuth, Altitude]. Then from LGN → V1, it is dramatically converted to the Log-polar coordinate system expressed by [Log(Eccentricity). Roll] (Horton and Hoyt 1991; Benson et al. 2014; Abdollahi et al. 2014). But surprisingly, it has not been investigated as a principle of face recognition in the human brain. **Figure 9-A** illustrates this frame translation; The left side is the 2D Linear coordinate, [Yaw, Pitch] = [Azimuth, Altitude], on the retina that faithfully represents the external space. The right side shows the observed Log-polar coordinate system at V1, [Log(Eccentricity). Roll], where the visual cortex V1, physiologically separated on the right and left occipital lobes, are superficially connected at the "Roll" angle = 0° (upward direction).

The fovea is the central range within ~4° of eccentricity, which is mapped to the left half of the Log-polar V1. We assume that two alpha brainwaves are traveling, one to the left and the other to the right, from the central vertical line of Eccentricity = 4°. Likewise, two more alpha brainwaves are traveling upward/downward from the central horizontal line of Roll = 0° (upward direction), corresponding to the counterclockwise and clockwise rotations.

Inspired by this peculiar Log-polar coordinate of V1, its application to computing algorithms for human face recognition has been investigated in the past by (Araujo and Dias 1996; Qi and Nakata 2006; Javier Traver and Bernardino 2010). **Figure 9-B** demonstrates the effect of the frame translation

on a simplified human face like an inverted triangle. A triangular human face on the left, mapped on the retina, is converted to the bizarre, distorted pattern on the Log-polar coordinate of V1 on the right. The lip is downward at ±180^o roll angle, the left eye is at -45^o, and the right eye is at +45^o. It is intriguing to realize that such a distorted image on V1 is what our brain generates as a starting point of face recognition. Nevertheless, this striking fact seems to have been overlooked in human vision research.

Why did evolution decide to twist shapes like human faces in such a peculiar fashion dramatically? We argue that there was a good reason, that is scaling and rotation for face recognition. The Log-polar coordinate of 2D retinotopy starts at LGN and continues to V1 →V2 → V3. But then at the PIT, the 2D retinotopy suddenly disappears (Abdollahi et al. 2014). This 2D →1D conversion strongly suggests that **NHT** (Neural Holographic Tomography) takes over at PIT and conducts the conversion from 2D (space) → 1D (space) + 1D (time). The generic mathematical treatment of the **NHT** was given in **Part II; Section 3**, using a 2D toy model. The same concept can be applied here. Instead of X-neurons and Y-neurons in the 2D toy model, one can consider "Log(Eccentricity)" neurons (green tiny circles) as X, and "Roll angle" neurons (yellow tiny circles) as Y, as shown in **Figure 9-A**.

Naturally, one can assume that the alpha brainwaves travel on V1/V2/V3 and PIT along with the horizontal and vertical directions, resulting in holographic encoding onto these 1D linear neurons for Log(Eccentricity) and Roll angle. Recently, traveling alpha brainwaves were reported in this occipital region (Lozano-Soldevilla and VanRullen 2019), which strongly supports the holographic encoding of the space.

These four lines of 1D neurons in **Figure 9-A** can be packaged into the lattice structure of **HAL**, as presented in **Part II; Section 5.3**. **Figure 10** shows possible packaging, which is named **2D Shape HAL**. Since the original **HAL** is designed to represent the 4D space-time and only 1D + 1D lines are required for the **2D shape HAL**, the remaining 2D, [Yaw, Pitch], will be reserved for the dorsal pathway and introduced in the following **Section 4**.

Human visual acuity appears to be better than 1,000 x 1000 = 10^6 pixels, at least within the fovea, with a total of ~10^8 photoreceptors (cones and rods). But this simple **HAL** alone cannot achieve such superior spatial resolution. The remedy will be proposed in **Part IV** by utilizing Discrete Fourier Transformation (DFT). A working example exists in the observed grid maps (Stensola et al. 2012). One can imagine that the same mechanism of DFT was developed for vision through evolution.

3.2 Scale/Rotation Invariance for 2D Shape Recognition

Now comes the original question; why does our primary visual cortex distort a perfect image on the retina so severely? It must be due to the specific feature of the Log-polar coordinate system that satisfies scale invariance and rotation invariance (Horton and Hoyt 1991; Benson et al. 2014; Abdollahi et al. 2014). Let's investigate the scale invariance first. To illustrate its effectiveness, **Figure 11** shows three cartoon faces with the identical shape of an inverted triangle. The middle **Figure 11-B** can be assumed to be a memorized face of ~8^o size. The top **Figure 11-A** is a factor two smaller (~4^o size), and the bottom **Figure 11-C** is a factor two larger (~16^o size).

As shown on the right side of the Log-polar coordinate, this scaling up/down by a factor of two becomes a simple horizontal parallel translation (of an order of 1 cm on the actual V1). Therefore, these faces in the Log-polar coordinate system are invariant under scaling. In other words, when we observe a smaller (or larger) face than the memorized size face by a factor of two, by shifting the phase of the horizontally traveling alpha brainwave by 1 cm to the left (or right), we can completely overlap the incoming face pattern and the memorized face on top of each other.

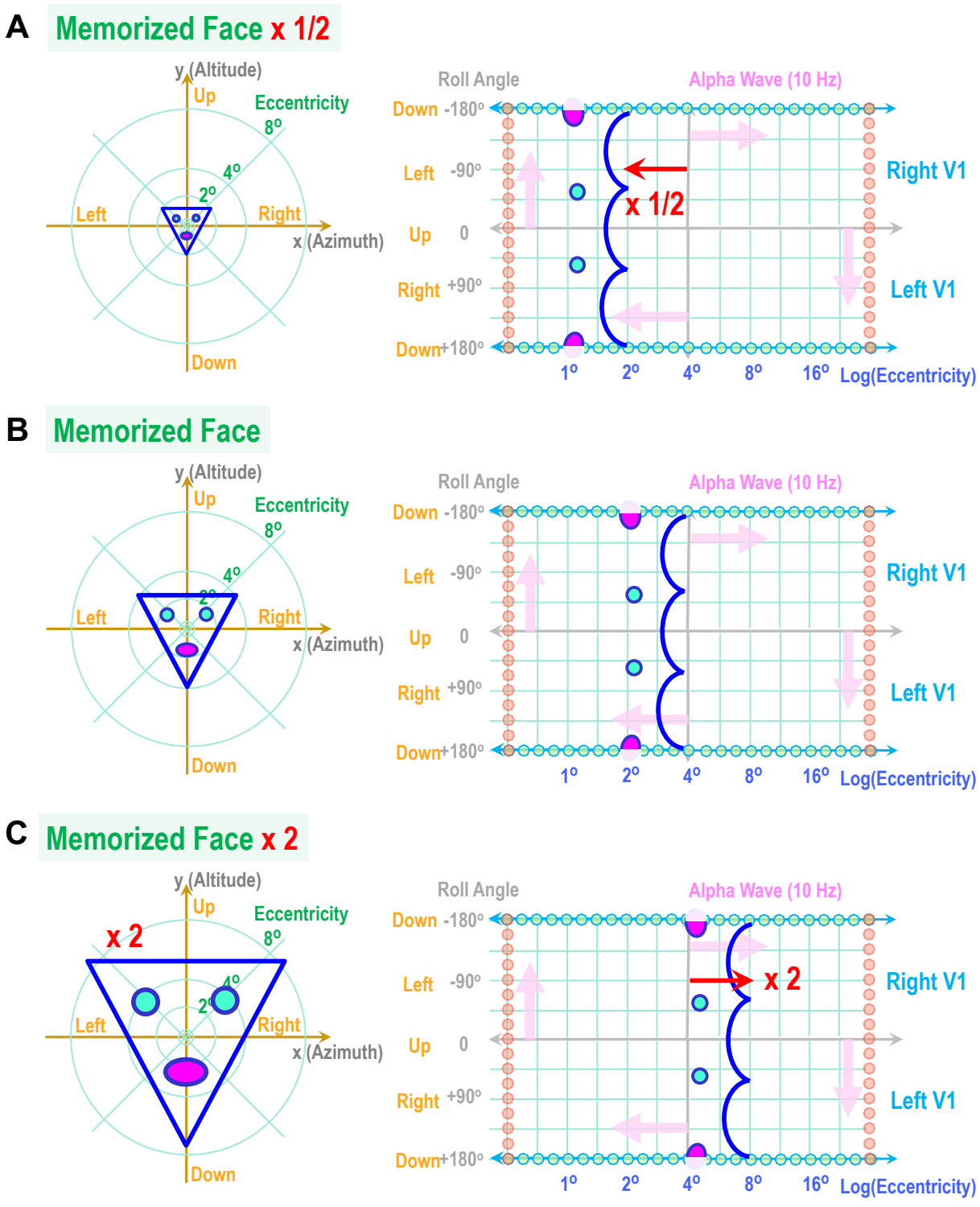

Figure 11. Representation of human faces with three different sizes in the ventral pathway with the Log-polar coordinate system. Here a simplified face (= an inverted triangle) is assumed. **(A)** is a half size of the memorized face of **(B)**, resulting in the linear translation of the image to the left on the Log-polar coordinate system. (B) is the memorized original size. **(C)** is twice larger than (B).

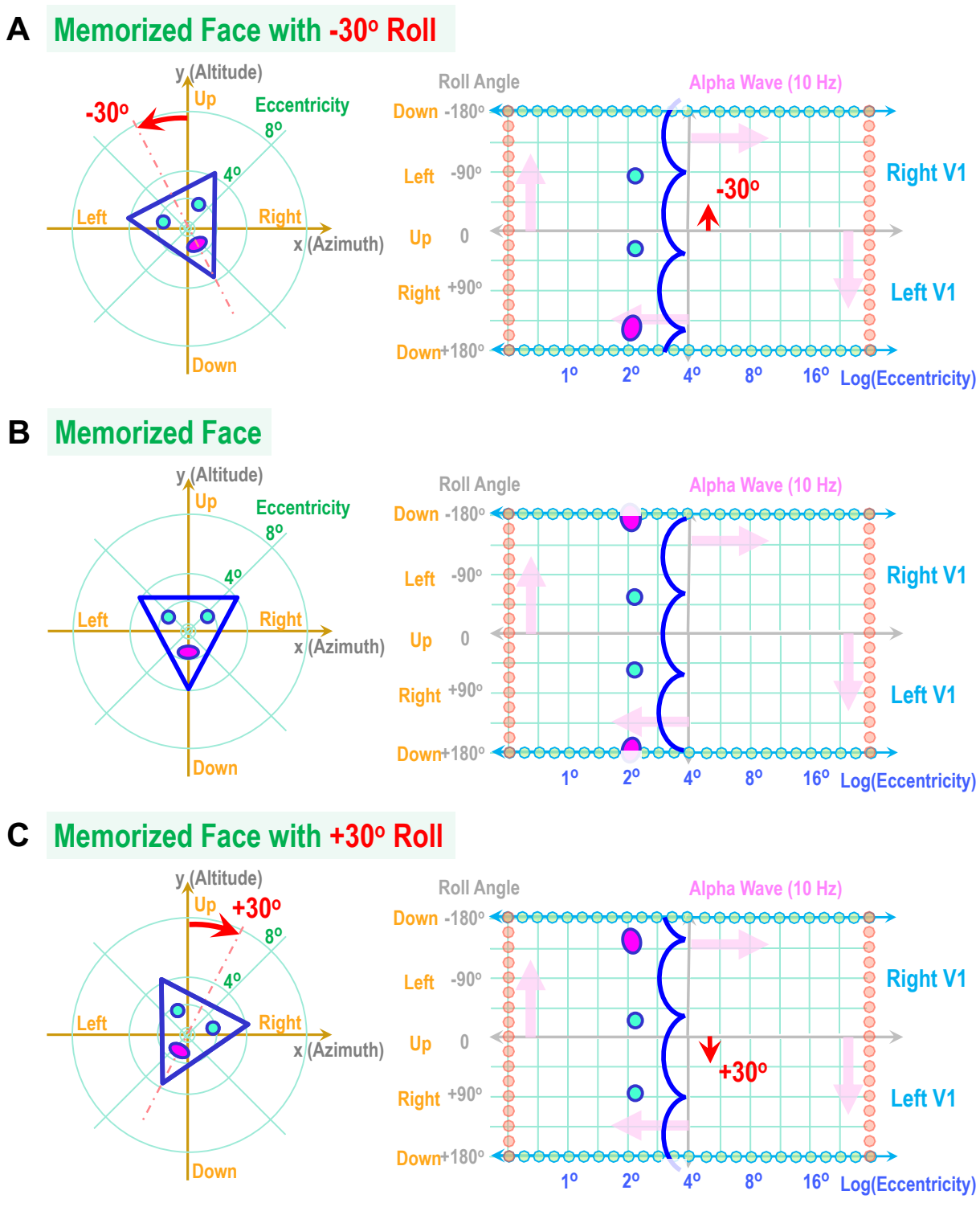

Figure 12. Representation of human faces with three different "Roll" rotations in the ventral pathway with the Log-polar coordinate system. **(A)** is 30° rotated left from the memorized face of (B), resulting in the linear translation of the image to the upward on the Log-polar coordinate system. **(B)** is the memorized original orientation. **(C)** is 30° rotated right from the memorized face.

This overlap happens in all the involved neurons as a time coincidence; thus, the whole process satisfies Hebbian plasticity, resulting in the brain-wide sensation of "conscious" awareness of this face as a semantic pattern. In other words, this is a moment of "aha, here is my wife" for example.

A classic reaction time experiment can prove this log-scaling effect in face recognition by showing different sizes of human faces on a TV screen. If a subject memorizes ~8° high faces (radius of ~4°) like **Figure 11-B**, then the exact size of the face gives the fastest reaction time. In contrast, when the face is a factor two smaller like **Figure 11-A**, or a factor two larger like **Figure 11-C**, the reaction time would become slower linearly by the scaling factor. We conducted this experiment and obtained a compelling result, which will be given in **Section 7**.

Similarly, we can consider rotation invariance under the "Roll" rotation shown in **Figure 12**. The middle **Figure 12-B** can be assumed to be a memorized face with no rotation. The top **Figure 12-A** is rotated anti-clockwise by 30°, and the bottom **Figure 12-C** is rotated clockwise by 30°. Like the case of scaling, these ±30° of rotation become up/down by 1 cm on the Log-polar V1. It means that when we observe a tilted face with ±30°, by shifting the phase of vertically traveling alpha wave by ±1 cm, we can completely overlap the incoming face pattern and the memorized face on top of each other. In summary, the log-polar coordinate of the V1/V2/V3 → Ventral pathway is powerful machinery to perform scaling and rotation for prompt face recognition.

We also conducted the reaction time experiment by rotation faces in 3D. We obtained the expected linear relation between the rotating roll angle and the delay of the reaction time, which will be covered in **Section 7**.

At what stage of the Ventral pathway are these scaling and rotation transformations happening? Since the VTC is known to recognize semantic shape, the scaling/rotation must occur before the VTC at somewhere between PIT → CIT → AIT. As shown in **Figure 8-A**, PIT receives retinotopic signals directly from V4 and converts the 2D map to holographic tomography by the **NHT** by alpha brainwave. So, it must be the object-centric frame. PIT also receives the retinotopic "linear" image from LIP/MIP after covert attention (see **Section 4.4**). We speculate that PIT could be the site of the object-centric Linear-polar coordinate system. Then at CIT, it returns to the Log-polar coordinate system. (Without covert attention, the bottom-up sensory signal from V4 bypasses PIT and directly goes to CIT.) Finally, at AIT, scaling and rotation translations must take place. AIT mutually communicates with the cortical aria 7a to send back and forth the shape of the object in the Cartesian form, but it also conducts proper scaling up/down to match the memorized shape at CTC.

3.3 Evolutionary Aspect: Insects, Rodents, Birds, and Primates

If the Log-polar coordinate system is so efficient at recognizing various sizes or orientations of complex shapes like faces, then one may speculate that it may be a generic feature of the primary visual cortex of any animal. The answer is no. Recent studies show that the mouse visual cortex is Cartesian (Garrett et al. 2014). [*To be exact it is a Linear-polar by* [*Yaw, Pitch*]*, but it would be nearly identical to the Cartesian of* [*X, Y*].] Likewise, the visual wulst (similar to V1) of the zebra finch exhibits the Cartesian coordinate system, too (Bischof et al. 2016).

Needless to say, insect vision is based on the Cartesian system as well (Caves, Brandley, and Johnsen 2018; Nériec and Desplan 2016; Otsuna, Shinomiya, and Ito 2014; Apitz and Salecker 2014; Sato, Suzuki, and Nakai 2013; Kolodkin and Hiesinger 2017). It is intriguing to see that in monkeys the development of the Log-polar coordinate system has started, but to be exact, their V1 still appears somewhere between the Log-polar and the Cartesian coordinate systems (Vanduffel, Zhu, and Orban 2014).

There may be a good reason to stick with the Cartesian for these animals. Only with the Cartesian coordinate system, maintenance of allocentric space by linear translation is possible. Without the allocentric frame, animals cannot navigate the space between food locations and their nests. Clearly, from an evolutionary point of view, the Cartesian must have come first. But then, these animals would have difficulties in recognizing semantic shape.

What does it mean? Probably before the evolution of the higher primates, such as monkeys and apes, vertebrates needed to perform discrete scaling by means of the Discrete Fourie Translation (DFT), which is a part of the **HAL** expression of the visual cortex. With a continuously varied frequency of alpha brainwave, non-primate vertebrates could also effectively scale up/down at any size. Evolutionarily, this might be the reason that visual systems of any animal, including rodents, incorporated the DFT to improve visual acuity without going to the scale-invariant Log-polar. Ants

Nevertheless, the DFT-based scaling is far less efficient than Log-polar-based scaling. The fact that even monkeys have not obtained the exact Log-polar visual cortex may indicate that only human vision (and perhaps that of apes, though this has yet to be tested) might be specialized to recognize complex semantic shapes at any size and rotation, especially to identify and distinguish various human faces. We will explore the topic of face recognition further in **Part IV**. The human-specific Log-polar system could have also contributed to the development of written language, which will be addressed in **Part VI**.

3.4 Summary – Ventral Pathway for 2D Shape Recognition

Traditionally, visual signal processing is considered bottom-up processing. Staring from Retina → LGN → V1 → V2 → V3 →V4 and so on, the Receptive Field (RF) becomes larger and larger by connecting dots and lines. Then finally at VTC, the entire shape of a complex object, like a human face, could be constructed and recognized. But true visual perception turns out to be far more complex and distinctive than this conventional wisdom. In our daily life, a human face may appear at any location in external 3D space, at any size, and at any orientation in 3D. Without a proper transformation of the observed face in 3+1+3 = 7D space, we cannot compare the observed face with the memorized face.

The ventral pathway is designed to conduct 2D scaling and rotation in an elegant manner. The human's primary visual cortex has evolved to map the sensory image from the retina to the Log-polar coordinate system that satisfies scale invariance and rotation invariance. As a result, scaling and rotation can be performed by a linear translation at CIT → AIT by the simple phase shift of alpha brainwaves.

In the following **Section 4**, we will describe the detailed biological mechanism of depth perception by integrating both dorsal and ventral pathways. Then **Section 5** will explore the dorsal pathway which maintains the 3D allocentric frame. Especially we will focus on overt and covert attention. By combining all three Sections together, we can finally answer the three major mysteries of our vision below. (The third mystery is split into two steps: (3) and (4).)

1) How we recognize the different sizes and orientations of shapes (like human faces): **Section 3**
2) How we faithfully perceive 3D space with depth**: Section 4**
3) How we maintain stable visual perception regardless of saccades: **Section 5**
4) How we reconstruct and perceive stable allocentric 3D space: **Section 6**

4 Holographic Perception of 3D Space with Depth

4.1 Mystery and Past Studies of Depth Perception

In this **Section 4**, we shall deal with arguably the most profound mystery of human vision; how can we perceive external 3D space with depth? Our depth perception seems so natural that we often take it for granted, but one should keep in mind that, like a camera sensor, the retina only receives 2D flat images. Some may guess that depth perception must be due to stereoview by the two eyes. It is undoubtedly one of the causes, but even a single eye can generate vivid depth perception under certain conditions, as we will describe later.

Historically, the cognitive process of 3D vision has been speculated about and evaluated under numerous causes of depth perception and sensation, including the ones listed below:

1) Triangulation by stereo view (by slightly inward gazing.)
2) Focusing on an object (by adjusting the focal length of the eye lenses.)
3) Binocular disparity (due to slight misjudgment of distance, together with micro-saccades.)
4) Optical flow (caused by the host's forward body motion.)
5) Motion parallax (caused by the host's side motion to the left or right)
6) Scaling of familiar objects with known absolute sizes (such as human faces and cars)
7) A linear perspective of converging straight lines to an infinity point.
8) Elevation effect (i.e., vertically lower is usually nearby.)
9) Kinetic depth effect (from moving 3D objects.)
10) Occlusion (i.e., shadowing by overlapping objects)

Among these, only (1) and (3) require a binocular view. The rest can be based on a monocular view. Due to the complexity of so many causes, past studies have been mostly psychological speculations based on daily behavioral experiences. The overview and recent development of stereopsis can be found in (Cumming and DeAngelis 2001; Nityananda and Read 2017; Vishwanath 2014). Among the above list, binocular disparity has been investigated the most (Chauhan, Héjja-Brichard, and Cottereau 2020; Lappin and Craft 1997; Okajima 2004; Railo et al. 2018). Far fewer studies have focused on the monocular origins. Motion parallax was studied by (Rogers and Graham 1979; Kral 2003; Kim, Angelaki, and DeAngelis 2016; Wexler and van Boxtel 2005). A possible neural basis was advanced by (Adesnik et al. 2012; Uka and DeAngelis 2004, 2006).

Unfortunately, it is fair to say that none of these publications could explore the problem in sufficient depth to reveal the true biological and physical principles behind depth sensation. At this point, the problem seems obvious; without the fundamental principle of space-to-time conversion, these traditional bottom-up approaches starting from 2D retinotopy would never yield depth sensation. To be precise, such approaches would not even generate a visual perception of 2D images either, as we already discussed extensively.

In this **Section 4**, we will first categorize the above causes from the point of view of **MePMoS**. Then we review the holographic principle of depth perception by applying the new concepts of **NHT** and **HAL** universally, step by step.

4.2 Various Categories of Depth Perception and Sensation

Type	*Function*	Binocular / Monocular	Local / Global	Motion	Process	Brain-wave	Local Displacement Vectors		
							Retina	Human V1	Animals
Type A - Eye Motion									
	Triangulation	Binocular	Global	Eyes	Focusing	Theta, Beta	None		
	Accomodation, Focusing	Monocular							
	Depth of Field								
	Binocular Disparity	Binocular	Local		Bottom-up		Horizontal	Rotating	Horizontal
	Micro-saccade	Monocular							
Type B - Body Motion									
	Motion Parallax	Monocular	Global	Body	Bottom-up	Beta	Horizontal	Rotating	Horizontal
	Optical Flow						Radial	Horizontal	Radial
Type C - Scaling									
	Scaling of known objects	Monocular	Global	None	Top-down	Alpha	Radial	Horizontal	Impossible
	Linear Perspective								
Type D - Top-down									
	Elevation	Monocular	Global	None	Top-down	Alpha	Vertical	Radial	Vertical
	Kinetic Depth Effect		Local				Complex		
	Occultation								

Table 2. List of the various causes of depth perception under four categories: **Type A** – based on eye motion, **Type B** – based on body motion, **Type C** – Scaling by brainwaves, **Type D** - Top-down inference by high-level image processing. The table is further organized in eight columns: (1) Binocular or Monocular, (2) local (relative) vs. global (absolute) depth, (3) caused by motion or not, (4) Bottom-up or Top-down process, (5) Which brainwaves,, (6) Direction of the displacement vectors on the retina as horizontal or radial, (7) on Human's V1 (with the Log-polar coordinate), and (8) on the primary visual cortex of other animals.

Let us begin with the categorization of the above-listed ten causes considering **MePMoS**. Basically, any visual sensation must be associated with **Me**mory → **P**rediction → **Mo**tion. Depth perception is not an exception. Fundamentally it should be associated with motions of either eyes, head, or body. If not, then motion must be substituted internally by traveling brainwaves running virtually into the depth direction.

In **Table 2**, we have divided the ten causes into four categories based on underlining movements (or no movement): **Type A** - eye motion, **Type B** - body motion, **Type C** – scaling by alpha brainwaves, and **Type D** - top-down inferences from higher-level cognitive processes by brainwaves.

Type A is stereopsis based on eye movements, including triangulation and binocular disparity. This type requires a binocular view and fine-tuning of gaze vectors of both eyes. A triangulation process matches the images from the left and right eyes perfectly well on V1, which gives an estimate of the absolute distance to the focused object at the foveal center. At the same time, slight binocular disparity

provides the 3D shape of the object or nearby distribution of other objects in 3D. Micro-saccades are an integrated partner of vivid 3D vision of stable objects. Focusing by eye lenses also helps triangulation to measure the distance to the focal plane. Involuntary eye movement is controlled by the theta brainwave. We will explain Type A in **Sections 4.4** and **4.5.**

Type B is based on body motion, including motion parallax and optical flow. In this group, a monocular view is sufficient because the host's head/body motion induces the dynamic flow of the 2D image on the retina. The image flow is mapped onto V1, then the motion vector segments are remapped onto log-polar retinotopy to derive distance because there is a well-defined one-to-one mapping from the motion vector to distance. According to our model, head/body movements are controlled by the beta brainwave. We will examine Type B in **Section 4.6.**

Type C is based on global 2D image processing on the scale-invariant Log-polar coordinate system. A good example is the scaling of familiar objects with known absolute sizes (such as faces and cars), like the 2D face recognition described in **Section 3.2**. Another example is the linear perspective of converging lines to an infinity point. These are top-down processes by alpha brainwaves after the 2D shape is recognized. Thus, monocular vision is sufficient, and nothing (eyes/head/body) is required to move. We will explore Type C in **Section 4.7.**

Type D is also based on the top-down process, but with an even higher level of cognitive image processing of the global scene or 3D-shaped objects. Examples are the elevation effect (i.e., vertically lower is usually nearby), the kinetic depth effect (from moving 3D objects), and occlusion (i.e., shadowing by overlapping objects). All of these require the internalized prediction of 3D space and shape. Basically, it is a consistency check between prediction and sensing by a holographic 3D projection mapping – the essence of **MePMoS** and **NHT**. We will summarize Type D in **Section 4.8.**

To systematically distinguish these four types, **Table 2** includes the eight columns below.

1) Is it based on the binocular or monocular view?
2) Is it sensitive to local (relative) or global (absolute) depth?
3) Is depth sensation caused by motion or static line segments on V1?
4) Is the process bottom-up or top-down?
5) What are the brainwaves, the Theta (~5 Hz), Alpha (~10 Hz), or Beta (~20 Hz)?
6) Is the direction of the displacement vectors in the Cartesian coordinate system on the retina?
7) What about on Human V1-V4 by the Log-polar coordinate system? and
8) What about the coordinate systems on the primary visual cortex (or equivalent) of other animals by the 2D Linear coordinate by [Yaw, Pitch]?

In the following sub-sections, we will examine all four types extensively, one by one, starting from the binocular disparity. Through this process, all the above parameters in the eight columns will be elucidated by employing **NHT** and **HAL**.

4.3 The Holographic Origin of the Depth Perception

Let us review once again the fundamental physical principle of observation of a specific location in 3D. In geometrical argument, we would suppose that a location (*x, y, z*) can be defined by three quantities: distance *x* along the *x*-axis, *y* along the *y*-axis, and *z* along the *z*-axis. For convenience, we usually assume the Cartesian coordinate system, where all three axes are orthogonal to each other. However, according to Einstein, the observation of location (*x, y, z*) requires a physical signal messenger, such as light, to travel from (0, 0, 0) to (*x, y, z*), and the required time to travel gives the observed distance. Human visual perception of space should not be an exception.

This principle is, of course, applicable in 2D and 1D as well. For a 2D retinal image to be perceived, the location (*x, y*) must be observed by two times (Δt_x, Δt_y). Even in 1D, to perceive and measure the length of a horizontal bar, Δx, a signal with constant speed *v* must travel to convert Δx to $\Delta t = \Delta x/v$. And that is the function of traveling brainwaves; Δt is expressed by the phase of alpha brainwaves. A detailed account of this fundamental principle of space-to-time conversion has already been given in **Part I** and **Part II**.

In **Part II**, we formulated this principle as **Neural Holographic Tomography (NHT)**. A beauty of this theory is that 3D visual perception is a natural extension from 2D visual perception in the time domain. If the 2D retinotopic image must be converted to time sequence by brainwaves, then depth perception can be included straightforwardly by adding one more brainwave traveling along the third axis (= the depth axis). This can be arranged effectively by the concept of **Holographic Ring Attractor Lattice (HAL)**.

We briefly summarized the principle of **HAL**-based 3D vision in **Section 1.5**. In the **HAL** model, the external three dimensions are reduced and represented by three sets of linear neurons: 3D → 1D (space) + 2D (time), where the time axes are expressed by the phases of six traveling brainwaves. **Figures 4** and **5** illustrate **3D Vision HAL** that realizes this compactification. Thanks to this time expression of depth, we can visualize the 3D space "*out there*" at the exact real location in the external world. This is the true physical origin of 3D visual perception.

So far, so good. But unfortunately, although the above argument defines 3D vision by the holographic principle, it still cannot explain how the 2D retinotopic image creates depth information at each local 2D point. And we must admit that it is a daunting task because every single point in 2D, like 1000 x 1000 = 10^6 points, must acquire depth information independently in parallel. How can the **HAL** be filled up with such 10^6 quantities properly by a physical and biological process through the visual pathways? The remaining **Section 4** is 100% devoted to investigating this nontrivial parallel-processing mechanism.

In this regard, an excellent starting point is provided by recent studies in electrophysiology. Researchers have shown that the entrance of the dorsal pathway, MT, is responsible for representing depth by binocular parallax (Nadler, Angelaki, and DeAngelis 2008), as well as by motion parallax (Kim et al. 2016). They observed that retinotopic neurons in MT flash at higher spike rates at each pixel if the specific 2D location corresponds to a larger binocular disparity or motion parallax. However, unfortunately, the biological origin of such enhanced spike rates has not yet been explained in these papers.

To fill in this missing piece of the biological origin of depth perception, we must pay attention to the bottom-up sensory input to MT. The MT receives visual stimulation directly from V1, V2, and V3/V4, as shown in **Figure 8-A**. So, we speculate that the sensation of depth must originate from signal processing in the primary visual cortex from V1 to V4. The prominent features of V1 to V4 are as follows:

1) The human primary visual cortex shows a retinotopic 2D matrix by the Log-polar coordinate system (Horton and Hoyt 1991; Benson et al. 2014; Abdollahi et al. 2014).
2) They exhibit orientation columns and ocular dominance (Yacoub, Harel, and Ugurbil 2008).
3) The neurons in the orientation columns of V1 enhance spike rates when observing local line segments or moving objects (Hubel and Wiesel 1959, 1962, 1968, 1998).
4) The above spikes induce gamma brainwaves; higher spike rates shift the phase of the gamma brainwave more, like ~10 ms for 90° change (Gray and Singer 1989; Fries et al. 2001; Fries, Nikolić, and Singer 2007; Fries et al. 2008; Vinck et al. 2010; Brunet et al. 2013)

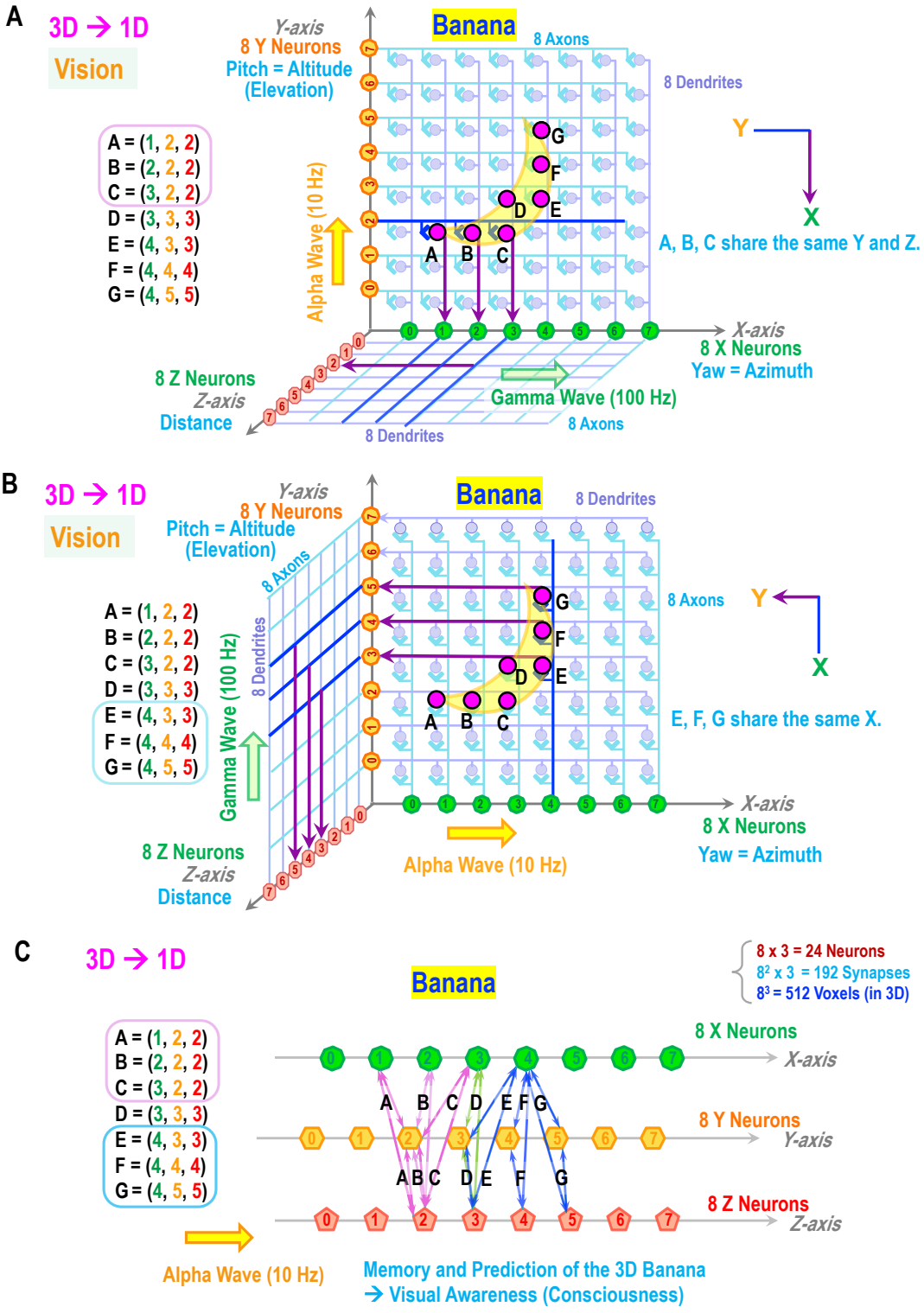

Figure 13. 3D toy model for 3D vision (duplicated from **Part II: Figure 9**). **(A)** and **(B)** show 8 x 8 x 3 = 192 synapses that record 8 x 8 x 8 = 512 locations in (X, Y, Z). **(C)** is equivalent to (A) and (B), but it demonstrates a compact arrangement of the memory unit established by the 7 (points) x 3 (combinations) x 2 (dual directions) = 42 synaptic connections.

All these facts play critical roles as essential individual pieces of the puzzle of depth perception. Roughly speaking, we hypothesize as follows. Firstly, at each retinotopic 2D location on V1-V3, either a local line segment or local motion vector is generated due to the distance to that location (corresponding to that direction in 3D space). For example, binocular disparity and linear perspective detect static local line segments. In contrast, motion parallax and optical flow generate local motion vectors dynamically. Secondly, these lines or motion segments selectively enhance the spike rates of specific neurons within the orientation columns. Thirdly, the increased spike rates induce the specific phase shifts of gamma brainwaves. Fourthly, these spike rates and gamma wave phases are transferred to MT for constructing the 3D image.

Finally, the puzzle is complete; Each retinotopic neuron in MT flashes with a specific rate with an assigned gamma phase, representing depth in that direction. And the time stamp given by the gamma phase is the local visual sensation of the specific depth at that location. Causality and locality are satisfied in 3D, as Einstein postulated.

Let's take a simple example of a banana in front of us as shown in **Figure 2**, which is duplicated from **Part II: Section 3.2 – Figure. 8**. This toy model is based on 8 x 8 = 64 pixels on the retina, but it can explain the basic principle of our 3D vision. (Also please note that three axes of [Yaw, Pitch, Distance] are substituted by the 3D Cartesian coordinate of [X, Y, Z] for simplicity.) **In Part II**, we already clarified the principle of how to encode the depth at each 2D pixel by gamma brainwaves in **Figure 9**, which is copied as **Figure 13**.

In **Figure 13-A**, the Y (upward) direction of the plane alpha brainwave is recorded by the eight X (right) neurons and the eight Z (forward) neurons as a function of time. For example, the position of Point A at (1, 2, 2) is recorded by Y = 2 neuron holographically with an X-phase of 1. But what about the depth information of Z = 2? Here, a different high-frequency gamma brainwave (as high as ~100 Hz) is required for the Z (depth) direction. It is because more than one point, like A, B, and C, flash together by the same phase of the horizontal alpha plane brainwave at Y= 2, but these three points must be distinguished by the independent timing for encoding possibly different depths (Z).

In the case of points A, B, and C, they happen to share the same Z, so the assignment of Z is not a problem. But let us consider the 3D encoding of points E, F, G by the alpha brainwave running to the right, shown in **Figure 13-B**. All three points, E, F, G will flash together by the same alpha plane brainwave at X = 3. Then, to distinguish the different depths, given by E = (4, 3, 3), F = (4, 4, 4), G = (4, 5, 5), the three different Y positions (3, 4, and 5) must be linked to the three different Z positions (3, 4, and 5) respectively. To achieve this, high-frequency gamma waves (f ~100 Hz) must run into the Z (depth) direction quickly. The enhancement of high-frequency gamma waves by visual stimulation has been observed on V1 (Gray and Singer 1989; Fries et al. 2001, 2007, 2008; Vinck et al. 2010; Brunet et al. 2013). We speculate that their observation corresponds to the bottom-up process of depth perception.

After all, the above processes convert the 3D banana shape, given by the seven points from A to G, initially detected by (X, Y, Z) = [Yaw, Pitch, Distance], to the seven sets of [X-phase, Y-phase, Z-phase] of the brainwaves. That is the complete expression of 3D visual perception by time; In principle, any 3D shape can be encoded and registered holographically and topographically. And this is the true origin of our 3D visual perception; we visually perceive the 3D space and shape by assigning the three times (= phases of brainwaves) at all the observed points. We argue that this is the only way to satisfy causality and locality in our visual pathway and construct the 3D image in front of us.

One should note that, through the above argument, the bottom-up process by high-frequency gamma brainwaves (~100 Hz) is assumed for holographic encoding of distance, whereas the initial retinotopic 2D by [Yaw, Pitch] is encoded by alpha brainwaves (~10 Hz). We assume that for conscious visual perception of 3D, the top-down reconstruction of 3D space is conducted by alpha brainwaves even for the third direction of the distance.

[*This assumption is supported by the reaction time experiment for different depths, which will be reported in* **Section 7.**]

Thanks to the static frequency-independent structure of the **HAL**, it is feasible that the recording and reading of distance utilize the different frequency bands: gamma for recording and alpha for reading. Later in **Section 6** and **Part IV**, we will show that the 3D vision system expressed by alpha waves is further converted to the 3D navigation system by theta waves, which has been extensively studied in rodents' hippocampal navigation system.

4.4 Depth Perception by Stereopsis – Triangulation and Binocular Disparity

We shall begin with **Type A** – the most widely accepted cause of depth perception – stereopsis by triangulation and binocular disparity. We have been told that having two eyes allows for stereoscopic vision. It is true that binocular view by two eyes seems to create a strong depth sensation, especially for nearby objects, but its underlying biological mechanism is still poorly understood.

There are two major factors that contribute to stereopsis. One is triangulation that establishes absolute distance, and the other is binocular disparity that estimates the local curved structure in 3D or relative distance to a nearby object. To illustrate the basic principle of each case, let us assume a spherical ball with a 3 cm radius like an apple or orange (instead of a banana), placed at 55 cm away in front of us, as shown in **Figure 14-A**. If our two eyes are focused on the ball correctly, assuming 6 cm separating between both eyes, both eyes must be pointed slightly inward by η = (6 cm /2) /55 cm ~ 0.055 rad ~ 3.1°. This is the origin of depth perception by triangulation based on **MePMoS**. The inward eye **Mo**tion generates the prediction of depth by **MePMo**. Followed by the predicted **S**ensation of the overlapping image of the ball by the left and right eyes, **MePMoS** is complete, and the distance to the ball is estimated. [It is worth noting that (Triangulation angle) = (Distance between two eyes) / (Distance to the object), thus Log(Triangulation angle) is linearly related to -Log(Distance). This linear relation will be explored later in conjunction with the Log-polar coordinate of V1.]

In contrast, binocular disparity is caused by misjudging the relative distance, say, 50 cm instead of the correct 55 cm away, causing a 5 cm misjudgment. Such an underestimate of distance by 5 cm would result in the observation of two slightly displaced balls as shown in this figure. In this case, the ball (with apparent diameter = 3°) is split into two images with the displacement of 0.3° in between. This phenomenon of dual images is called binocular disparity. As we discussed in **Section 2**, the left-eye image and right-eye image are transferred to LGN → V1, where the images are overlapped, mixed, and distorted on the Log-polar coordinate system, as illustrated in **Figure 14-B**. If we observe a perfect circle with a 3° radius, a straight vertical line would show up on the V1. But due to the binocular disparity, we observe slightly off-centered two balls by 0.3°. These two circles will show up as wiggled orange and green lines.

Let us consider the displacement vectors along the boundary of the ball given by many tiny gray arrows. The parallel gray arrows on the Cartesian coordinate system on the left side (on the retina) in **Figure 14-B** are mapped onto the gray arrows with a circular pattern on the Log-polar coordinate system on the right side (on V1). These tiny gray arrows (with 0.3° length) on V1 will trigger the ocular dominance neurons sandwiched by the orientation column (Yacoub et al. 2008). The critical point is that there is a one-to-one mapping between the displacement vectors on V1 (generated by the binocular disparity) and the miss-measured relative distance. In this case, all the displacements show 0.3° over 3° which is 10%.

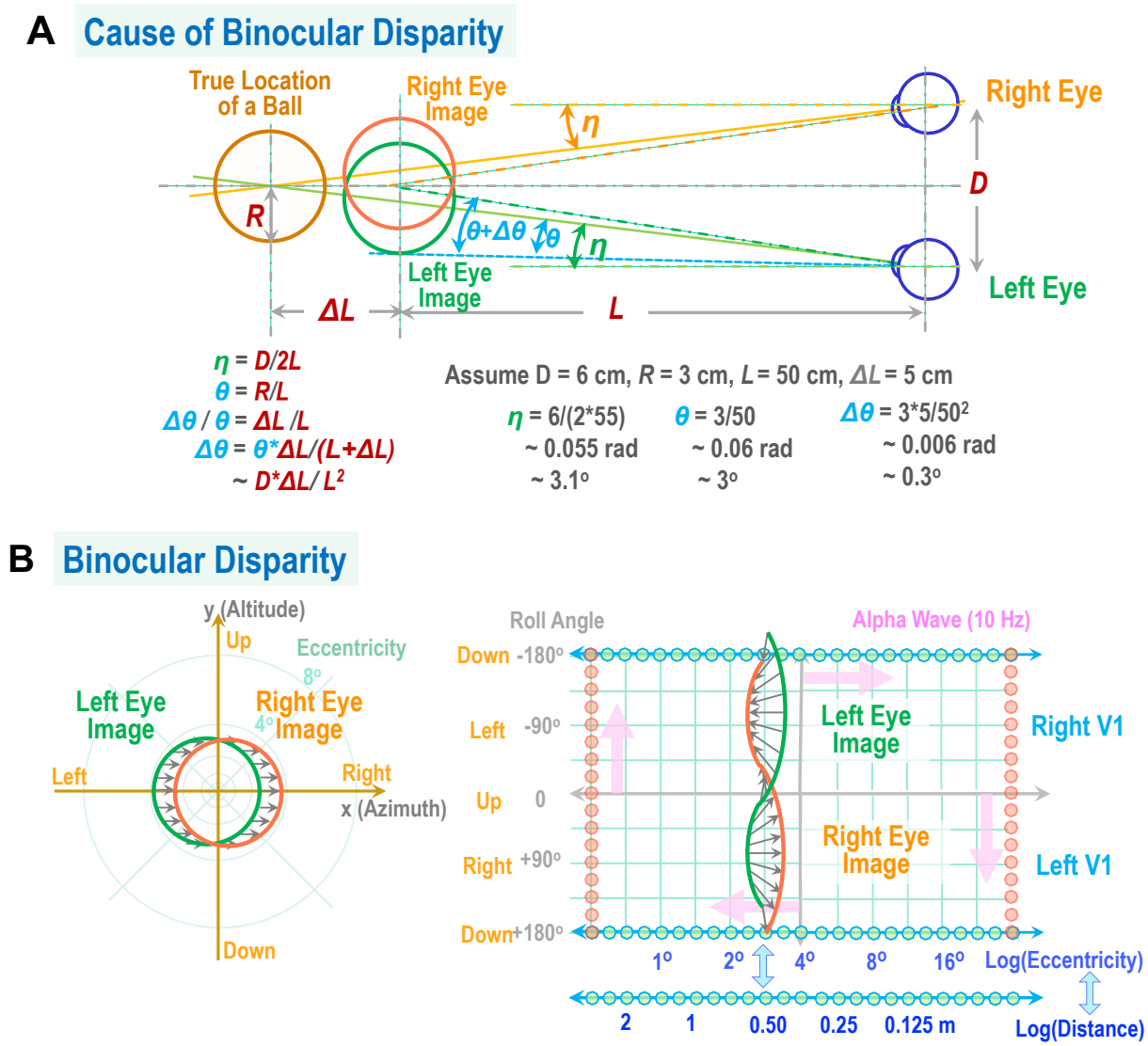

Figure 14. Depth perception by Binocular Disparity. **(A)** shows the cause of the binocular disparity. Misjudgment of the depth introduces double images, displaced by **Δθ**. (**B**) shows the effect of binocular disparity on V1, where the orientation column and the ocular dominance column work together to pick up tiny orientation vectors. **(C)** shows **3D Log-Polar HAL** by [Roll, Log(Eccentricity, Log(Distance)].

These 10% displacement vectors will enhance the spike rates accordingly in the orientation column (Hubel and Wiesel 1959, 1962, 1968, 1998). Then spike rates are converted to the phases of gamma brainwaves (Gray and Singer 1989; Fries et al. 2001, 2007, 2008; Vinck et al. 2010; Brunet et al. 2013). Consequently, all the neurons along the boundary of the ball encode the estimated distance, that is 50 cm x (1 + 0.3°/3°) = 55 cm. The enhanced spike rates and gamma phases are transferred to the MT as observed (Nadler et al. 2008). This is the biological origin of the depth sensation caused by the binocular disparity.

The dorsal pathway goes from MT → VIP/MST → LIP/MST → FEF. Through this process, the depth perception will be refined by incorporating many contributions other than stereopsis. After FEF the retinotopic image with depth information is converted to the holographic representation of **3D Linear-polar HAL** shown in **Figure 14-C**. At this stage, the 3D location of the boundary of the ball is represented by the three alpha wave phases including the depth direction. Initially, it should be the 3D Log-polar coordinates of [Roll Log(Eccentricity), Log(Distance)] consistent with the Ventral pathway. But then, it will be further transferred to 3D Linear-polar coordinates of [Yaw, Pitch, Distance], as described in the next **Section 6**.

Triangulation directly estimates the absolute distance by (Distance to the object) = (Distance between two eyes) / (Triangulation angle). In contrast, binocular disparity only provides the estimated relative displacement from the correct distance. The effect of binocular disparity is scale-invariant on the Log-polar coordinate system of V1; If the ball is located at twice the distance at 1.1 m, the ball diameter becomes half, and binocular disparity becomes half. As a result, the sensation of the 3D curved shape (surface) is accurately represented in our visual perception of the nearly perfect 3D Cartesian space. Thanks to the Log-polar nature of our visual cortex, we can maintain the proper 3D shape of an object regardless of the distance to the object. We will revisit this argument in **Section 4.6**.

Lastly, the eyes' lenses are automatically focused on the object of interest at the foveal center. The eye muscles for lens focusing should also produce a corollary discharge, which can predict the distance to the object, following the principle **MePMoS**. This situation is similar to triangulation. Basically, a perfect focusing onto an object at a certain distance forces eyes' muscles to tilt inward perfectly and relax/tighten muscles to fine-tune the thickness of the lenses. All of the rearrangements of muscles should be able to generate the corollary discharges to re-assign the proper phases of alpha brainwaves to the focused distance. [*Our reaction time experiments show that the focused point in 3D resets the alpha brainwaves to the zeros along [Yaw, Pitch, Distance] axes. See **Section 7**.*]

4.5 Micro-Saccades by MePMoS for Static Images

Micro-saccades are a somewhat peculiar phenomenon; even if we focus on one spot, our eyes are unconsciously slightly moving horizontally by an order of 0.1 – 0.5 degrees, periodically a few times per second. Past researches and reviews of micro-saccades can be found at (Laubrock, Engbert, and Kliegl 2005; Engbert 2006; Melloni et al. 2009; Martinez-Conde et al. 2013).

According to our theory of **MePMoS**, it is an essential partner of binocular disparity for vivid visual 3D perception. Let us assume ~0.3° of horizontal micro-saccade to the right, while we are watching a spherical ball ~50 cm away like in the previous example of the binocular view. The retinotopic images on both eyes are slightly shifted from the left (before a microsaccade) to the right (after) as shown in **Figure 15-A**. Remarkably, these two-ring patterns are identical to the case of binocular disparity in **Figure 14-B**. Such a similarity suggests that both micro-saccades and binocular disparity, share the unified principle to enhance visual perception of an object's 3D shape at the foveal center.

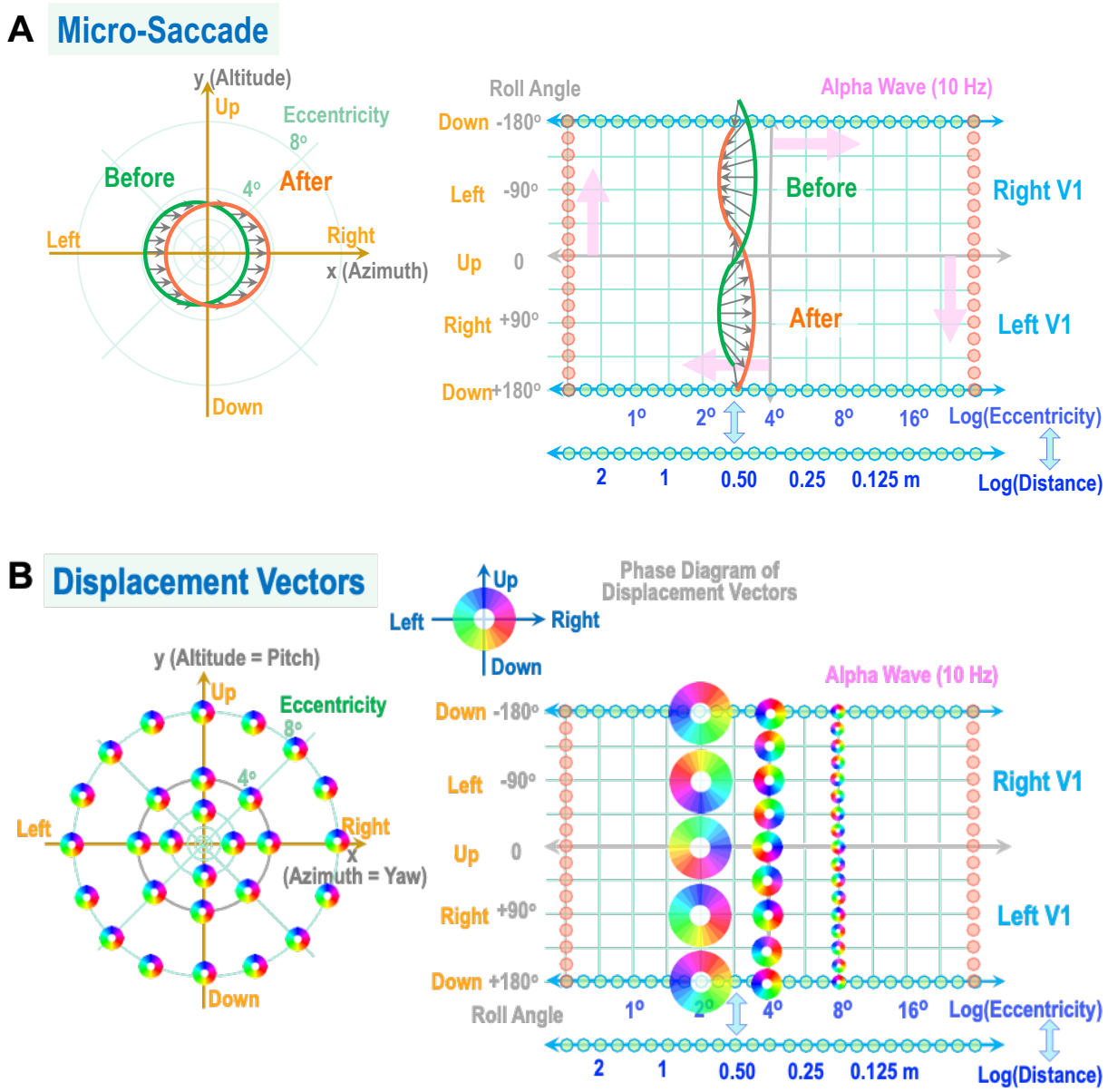

Figure 15. **(A)** Displacement vectors by a microsaccade. The pattern is remarkably identical to the depth perception by binocular disparity. **(B)** Phase diagram of the displacement vectors. This shows the basic principle of depth perception at any local 2D location in [Yaw, Pitch]. These displacement vectors are converted to the spike rates, then to the gamma phase, representing the depth at the given 2D point.

To appreciate the fundamental function of micro-saccades in the **MePMoS** model, let's start from light striking the retina. Once light hits photoreceptors (cones and rods), signals are immediately processed by horizontal cells for contrast enhancement in space, followed by bipolar cells to enhance transient signals in time. Thus, images are immediately processed by differentiation in space and time. That is why a stationary image would fade away in a few seconds without saccades to recover it. From an evolutionary point of view, micro-saccades optimize the sensitivity to moving objects like prey and

predators. On the other hand, this process eliminates images of stationary objects. [*A good example is a strategy of a silverfish; it suddenly stops and freezes on a floor when we find and try to kill it*.]

Micro-saccades evolved to recover fading stationary images, such as a stopped silverfish, efficiently in a predictive manner. By shifting the image to the right (or left) by a fraction of a degree, the sensation of a moving image can penetrate through the horizontal and bipolar cells, then it will land on V1 as shown in Figure **15-A**. Since this image shift is caused by a small saccadic eye movement in a well-controlled manner, the faithful effect copy (= corollary discharge) can predict the exact location of the new image at the exact timing after a micro-saccade. This is a perfect example of the power of **MePMoS**. Only a stationary image perceived just before the saccade will be enhanced and rewarded at V1. As a result, the neurons in the orientation columns that detect the direction vectors, given by the rotating tiny arrows (~0.3^{o} length) in **Figure 15-A**, enhance the spike rates and then generate the phase shift of gamma brainwaves. But unlike the case of the binocular disparity in **Figure 14-B**, the neurons involved in ocular dominance are not involved. Thus, no extra depth sensation is generated.

In a way, a micro-saccade acts as a calibration to ensure a sharp image with good contrast on the focal plane. Combined with binocular disparity, a 3D-shaped (curved) object, like a banana or an apple, is accurately reconstructed from the focal plane to slightly off-focused distances, like tracing the 3D-curved surface of an object.

Interestingly, on the Log-polar coordinate of V1, the displacement vectors by the micro-saccade are dramatically enlarged toward the foveal center. **Figure 15-B** illustrates such an effect by the phase diagram of 0.3^{o} displacement vectors; a micro-saccade to the right is presented by the red-color displacement vector. On the Log-polar coordinate at V1, displacement vectors are rotating and enlarged towards the foveal center. These patterns and sizes of vectors are fully predictable for each micro-saccade as well as binocular disparity. It tells that human vision has superior depth perception in the fovea in addition to superior 2D visual acuity.

4.6 Depth Perception by Body Motion – Optical Flow and Motion Parallax

Let us move on to **Type B**. Even from a monocular view, we can still obtain a strong depth sensation by physically shifting the eyes' location by a body motion in the allocentric frame. In the case of saccades, the eye's gaze vectors are re-orienting while maintaining the centers of eyes fixed. In contrast, in Type B, the centers of eyes are physically moving in 3D space. There are two examples: Motion parallax and optical flow, shown **in Figure 16**.

Figure 16-A displays the effect of motion parallax caused by a body motion to the right. Here, the central view (4^{o} from the foveal center) is ~4 m away, resulting in tiny displacement vectors to the left, opposite to the body motion. In contrast, 8^{o} from the foveal center observes ~2 m away objects, so it produces much larger displacement vectors to the left. But once these are mapped to the Log-polar V1, similar patterns of rotating orientation vectors are sensed by the orientation column.

In contrast, **Figure 16-B** shows the effect of optical flow caused by a forward body motion. Here, the central view (4^{o} from the foveal center) is ~ 4 m which produces tiny displacement vectors radially outward. In contrast, at 8^{o} from the foveal center observes ~2 m away objects, which produces much larger displacement vectors to the right. Once these are mapped to the Log-polar V1, similar horizontal patterns are sensed by the orientation column.

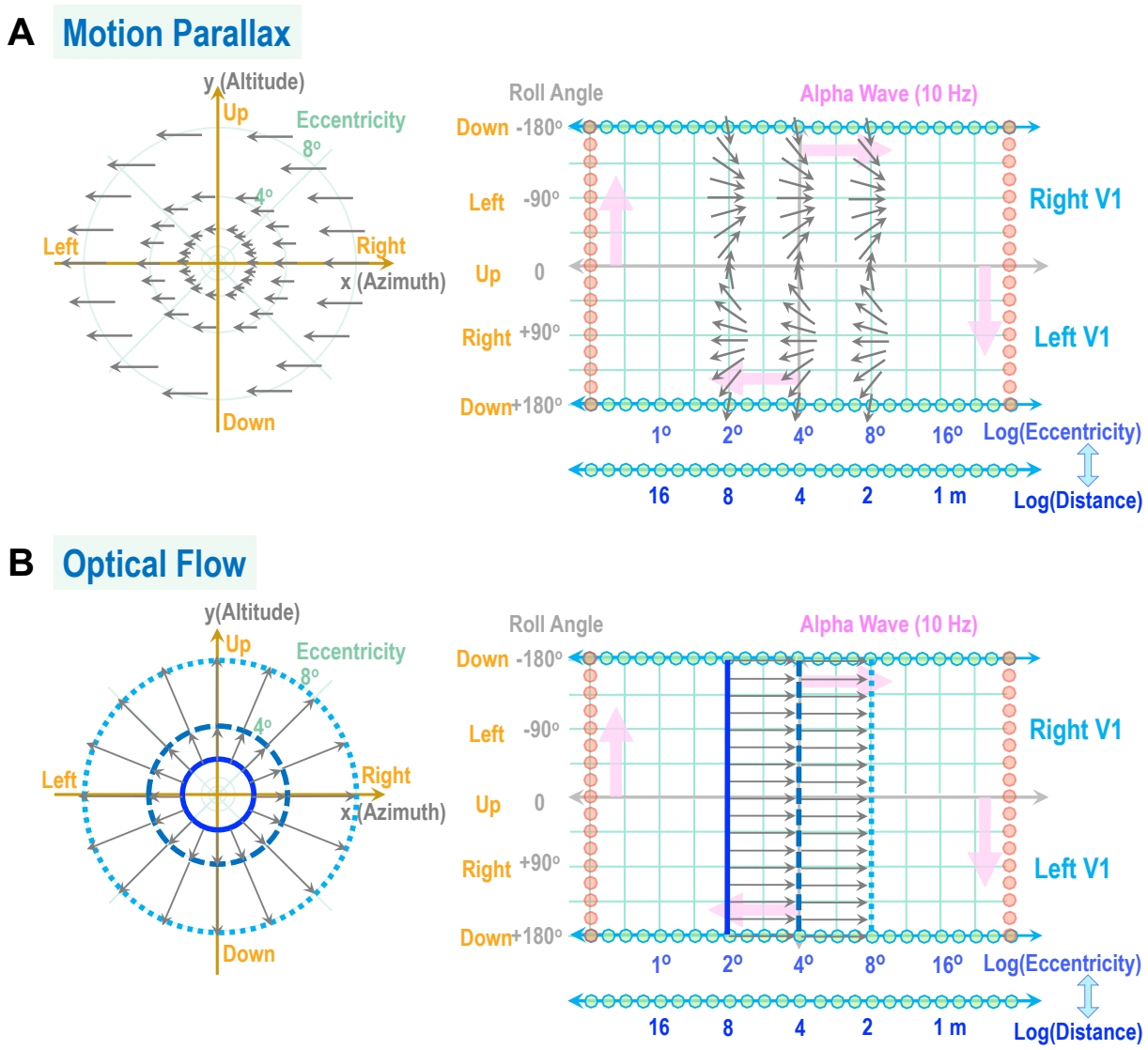

Figure 16. 3D perception by Motion Parallax and Optical Flow. **(A)** shows the effect of motion parallax, where the centra view (2.5° from the foveal center) is ~8 m away, resulting in tiny displacement vectors to the left. In contrast, 10° from the foveal center observes ~2 m away objects, resulting in much larger displacement vectors. Once these are mapped to the Log-polar V1, similar patterns of rotating orientation vectors are sensed by the orientation column. **(B)** shows the effect of optical flow, where the central view (2.5° from the foveal center) is ~ 8 m that it produces tiny displacement vectors radially outward. In contrast, 10° from the foveal center observes ~2 m away objects. Once these are mapped to the Log-polar V1, similar horizontal patterns are sensed by the orientation column.

In both cases, these processes are governed by **MePMoS**. For a given body movement in 3D space, either to the left/right, up/down, or forward/backward, an external object at the specific allocentric 3D location exhibits the specific displacement vectors that are highly predictable. The directions of the displacement vector are opposite to the body motion, and their magnitudes are inversely proportional to the distance.

The predicted displacement vectors (by the corollary discharge) are mapped onto V1 in time just before the new image (after the body motion) arrives. Overlying the new image on top of the predicted image will confirm the distance at every single neuron on the 2D retinotopic image, in real-time with parallel processing. This 2D-distributed information of depth is represented by spike rates and gamma wave phases by the corresponding neurons.

Type B depth perception is the most primitive and ancient mechanism among the four types. For example, recent research showed that insects can observe motion parallax and optical flow to maintain allocentric direction while flying at least within a 2D flat plane (Apitz & Salecker, 2014; Caves et al., 2018; Kolodkin & Hiesinger, 2017; Nériec & Desplan, 2016; Otsuna et al., 2014; Sato et al., 2013; Mauss et al. 2015). We speculate that insects can utilize 2D vision to generate depth perception by motion parallax caused by the upward/downward motion, too.

One should note a specific feature of insects' sensing to optical flow. While they are flying in the open air, they often experience airflow (= wind), which drifts their bodies along with the direction of the airflow. Such body drift is not caused by their own efforts by muscles, thus a corollary discharge is not generated. According to our model of **MePMoS**, they cannot predict these displacement vectors; thus, optical flow is not rewarded. Nevertheless, it is known that insects can take optical flow as an indication of its own motion (opposite to the flow) within the allocentric 3D environment.

The same can be said when we are driving a car. Optical flow is caused not by our body motion generated by muscles but by the car's motion, while we are sitting in the driver's seat (without any muscle motion, thus no corollary discharge.) Why is it possible to maintain the perception of a stable allocentric space, and sense the car's forward locomotion like a GPS system?

We speculate that a similar mechanism to insect navigation with optical flow plus displacement by wind must have been inherited by our navigation system. In **Section 6.4**, after we discuss how we maintain allocentric space regardless of eyes/head/body motions, we will revisit this concern and will provide a possible solution.

4.7 Depth Perception by Scaling and Linear Perspective

In the case of **Type C**, human vision seems to be able to estimate depth by scaling 2D images, even with a monocular view unlike triangulation or binocular disparity, and without body motion and thus without motion parallax or optical flow. Two good examples are scaling of familiar shapes and linear perspective, shown in **Figures 17-A** and **18-A**.

In our daily life, we often experience quite a strong sensation of depth when we observe numerous people of different sizes. For example, when we give a presentation on a platform in front of a large audience, we immediately sense the distances to persons in the audience by simply watching different sizes of their faces. Or when we drive a car, we sense the distances to all the incoming cars simply by their apparent sizes. Without such prompt depth sensation, driving a car would be dangerous. **Figure 17-A** demonstrates how we estimate distances from the scaling of familiar shapes like human faces. We all know that a typical human face is ~20 cm high. Thus, if it is located 1.4 m in front of us, it generates a pattern of $\pm 4^{o}$ on the retina. But if the same face is placed only at 70 cm away, the apparent size is doubled to $\pm 8^{o}$ on the retina. On the contrary, if it is 2.8 m away, it shows up like $\pm 2^{o}$. Once these faces (inverted triangles) are remapped onto the Log-polar coordinate system of V1, the identical patterns show up with horizontal parallel shifts as shown in **Figure 17-B**.

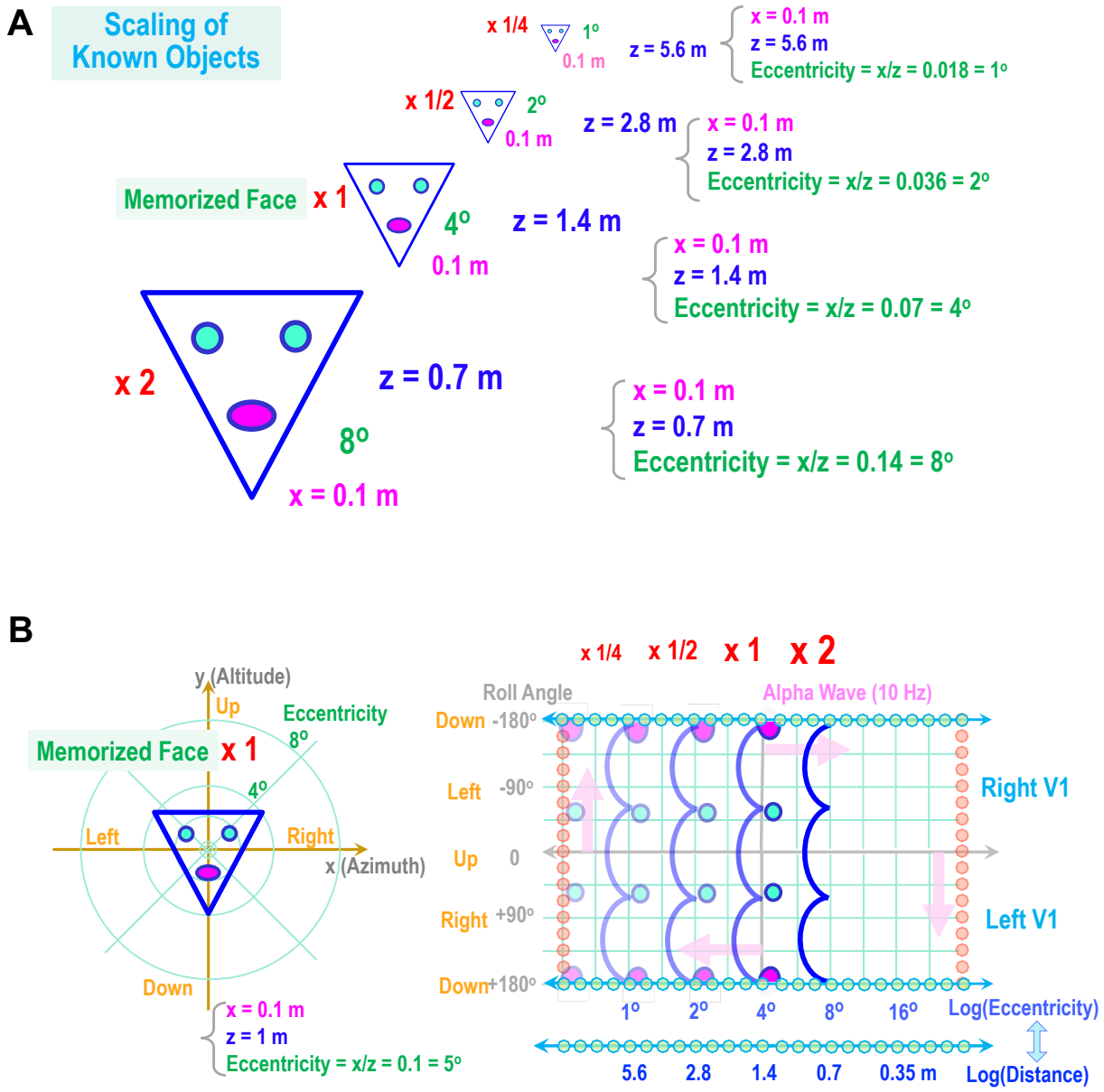

Figure 17. Depth perception by scaling objects of a known size such as a human face. **(A)** Illustration of depth perception of watching human faces at four different distances: z = 0.7, 1.4, 2.8, and 5.6 m. The apparent sizes of these faces will be ±8°, 4°, 2°, 1° in eccentricity respectively. **(B)** The scaling factor gives you depth perception by the linear image shift on the Log-polar coordinate of V1.

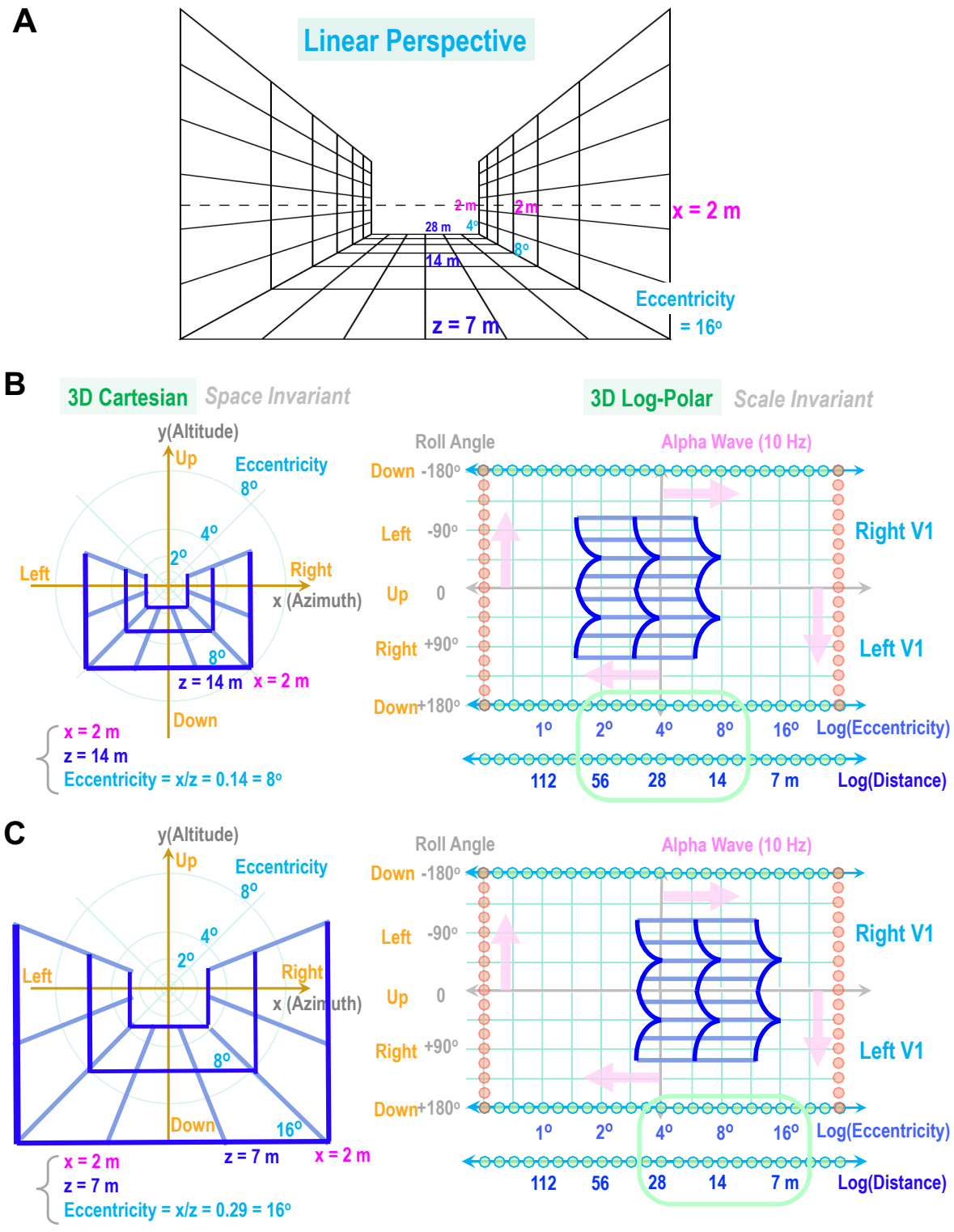

Figure 18. 3D perception by linear perspective. **(A)** is an example of linear perspective, produced by multiple lines on the floor and the left/right walls (at x = ±2 m from the center), converging to the foveal center. Eccentricity is given by z (distance) / x (width). **(B)** and **(C)** show the mapping of (A) to the Log-polar V1, corresponding to our experience when driving along straight road. The converging lines on the Cartesian system become parallel lines in the Log-polar system, and the observed optical flow runs from the left to the right at a constant speed.

Section 3.2 already introduced scale invariance by the repeated identical pattern shown in **Figure 11**; every factor of two enlargement (or reduction) of the face shifts the original mage by ~1 cm on V1 to the right (or to the left). At the left-bottom of **Figure17-B**, the Log(Distance) axis is added in parallel to the Log(Eccentricity) axis. Taking the simple relation of (Eccentricity) = (20 cm height of the face) / (Distance), the Log(Eccentricity) axis can be directly mapped onto the axis of -Log(Distance). Such hard-wired parallel wiring is sufficient to instantly convert the observed face size (= eccentricity) to the estimated distance. As a result, the induced sensation of distances to different sizes of human faces must be prompt, accurate, and vivid thanks to the extremely robust one-to-one hard-wired mapping.

Remarkably, this depth perception by such scaling can be achieved even by monocular vision, and even without any movement of eyes, head, or body. Essentially, it is purely based on the top-down prediction initiated from the past memory of a typical human face, conducted by the traveling alpha brainwaves on the Log-polar coordinate system, another example of **MePMoS**.

What about the linear perspective shown in **Figure 18-A**? Here again, strong depth sensation is induced due to the converging lines to the foveal center, as far as we focus the eyes to the center of converging lines. We always experience this effect when we are driving a car on a straight highway or walking on a long straight street surrounded by tall square buildings on both sides. To be exact, we do not have to drive or walk; Just a static flat image of **Figure 18-A** is enough. It is worth noting that we experience the same strong depth sensation regardless of binocular or monocular views, but once we shift the focus point from the center to the periphery, we lose this sensation. So, it is easy to conclude that the pattern of radially converging straight lines towards the foveal center is the cause of depth perception by linear perspective.

Here, we assume a street with the horizontal width of ±2 m (4 m wide) starting at z (distance) = 7 m, which results in the eccentricity = (2 m)/(7 m) ~ 0.3 rad ~ 16^{o}. Both sides are surrounded by vertical walls. The linear perspective image on the retina is remapped to the repetitive pattern on the Log-polar coordinate system as shown in **Figures18-B** and **C**, where the converging lines on the retina become horizontal parallel lines. This remapping is similar to the case of scaling faces in **Figure 17**.

Once again, Log(Eccentricity) axes can be directly remapped by hardwired connections onto -Log(Distance) axis. Please note that Eccentricity = 8^{o} corresponds to Distance = 14 m in this case, whereas in the case of scaling faces, 8^{o} corresponds to 0.7 m. In other words, the conversion factor from eccentricity to distance depends on the expected absolute size of the object of interest. That is why this must be a purely top-down mechanism based on memory and prediction.

The critical feature of Type C (with no motion) is the utilization of the scale-invariant Log-polar coordinate system on V1, [Roll, Log(Eccentricity), which only the human has. All other animals have Linear coordinates of [Yaw, Pitch]. (Please refer to **Section 3.3.**)

Therefore, Type C of depth perception is very specific to human vision. As it is based on the Log(Distance), Type C has an extremely long lever arm (sensitivity) to the long-distance among the three types. In contrast, Type A by binocular view has the shortest lever arm. One can conclude that humans sense longer distances even without any motion compared to any other animal species. Type C effectively gives the best predictive power for the largest space-time in the environment, which must have helped for the survival of our ancestors. Perhaps it was the evolutionary shift to open grassland environments and to a hunting and gathering subsistence pattern that put selective pressure on long-distance vision in humans.

4.8 Depth Sensation by Top-down Image Processing

Lastly, **Type D** includes all other indirect processes of depth perception by higher-level cognitive inferences. It is fundamentally the top-down processing of the global scene or 3D-shaped objects. Three good examples are:

1) The elevation effect: vertically lower is usually nearby.
2) The kinetic depth effect: a moving/rotating 3D object generates a predicted dynamic 2D pattern.
3) Occlusion: A frontal object shadows an object behind it.

All of these require a prior internalized prediction of 3D space and shape. Basically, it is a consistency check between prediction and sensing by a holographic 3D projection mapping – the essence of **MePMoS** and **NHT**. In other words, as far as the observed sensory 2D retinotopic image is consistent with the predicted 3D shape and location, the projected 3D internal signal is rewarded and consistent with depth sensation.

4.9 Summary – Grand Unification of Depth Perception

In this section, by applying the concepts of **MePMoS**, **NHT**, and **HAL** altogether, we were able to elucidate several different causes of depth perception in a truly unified manner. The most critical concept is that the visual sensation of depth is generated by the brain internally by assigning the predicted distance by means of alpha brainwaves' phases like a hologram, point by point in the 2D visual field of [Yaw, Pitch]. It is like a 3D projection mapping into empty space.

To be precise, this concept of holographic depth perception seamlessly integrates several products from evolution listed below.

1) Remapping of retinal images onto the primary visual cortex with the scale-invariant Log-polar coordinate system (which was likely the latest evolutionary step toward contemporary humans.)
2) The orientation column and ocular dominance on V1 which detect displacement vectors, corresponding to the depth at that location (direction).
3) Enhancement of local spike rates of neurons in the orientation column, representing the direction and magnitude of local displacement vectors (from V1 → V4).
4) Conversion from spike rates to phase shifts of gamma brainwaves.
5) Creation of the scale-invariant **3D Log-polar HAL** along the ventral pathway.
6) Finally, frame transformation
    a. from the **3D Log-polar HAL** in [Roll, Log(Eccentricity), Log(Distance)]
    b. to the space-invariant **3D Linear-polar HAL** in [Yaw, Pitch, Distance].

The formation of 3D visual perception by the **3D Linear-polar HAL** is truly an integrated process between the ventral and dorsal pathways in both bottom-up (by local gamma brainwaves) and top-down (by global alpha brainwaves) processing. This new concept can be regarded as the **Grand Unification** of depth perception. Knowing this sophistication and unification, one could argue that depth perception in human vision may be the most remarkable achievement that occupies nearly half of our brain.

Based on the origin of depth sensation, we can split the causes to **Types A**, **B, C,** and **D** by binocular vision, body motion, scaling, and top-down processing respectively as given in **Table 2**. **Type A** is stereopsis including triangulation and binocular disparity. This type requires binocular vision and fine-tuning of gaze vectors of both eyes. Focusing by the eyes' lenses also helps triangulation to measure the distance to the focal plane. **Type B** includes motion parallax and optical flow. In Type

B, the retinotopic image mapped onto Log-polar V1 receives motion vectors, and they derive the distance because there is a well-defined one-to-one mapping from the motion vector to distance.

Type C is based on global visual image processing of 2D shape. This is a top-down process after the 2D image is recognized. A good example is the scaling of familiar objects with known absolute sizes (such as faces and cars). Another example is the linear perspective of converging lines to an infinity point. This is a top-down process controlled by alpha brainwaves. **Type D** is also a top-down, but an even higher-level cognitive process based on predicted 3D shape/location. Examples include the elevation effect, the kinetic depth effect, and occlusion.

Among these, binocular vision is only required for **Type A**. Contrary to our conventional wisdom, depth perception and sensation are primarily caused by the monocular vision by **Types B**, **C**, and **D**, especially for longer distances (>> 10 m away). Within these monocular origins, the strongest depth perception can be generated by **Type B** with body motion. But even with a monocular view without any motion, our visual system is intelligent and creative to generate some depth perception based on top-down predictions by **Types C** and **D**.

The above basic principles can explain some intriguing experiences in daily life. A few decades ago, 3D televisions (with special polarized eyeglasses) became fashionable but eventually faded away. Why? It was designed based on the assumption that depth perception was solely caused by binocular disparity, which was too naïve. Our depth perception is far more sophisticated; monocular vision alone can produce vivid depth sensation.

Creating depth perception by a flat TV is non-trivial because Type A generates a flat 2D image on a TV by the binocular view (as it should be), while Types B-D tries to establish the proper depth sensation by the monocular view. The inconsistency between the binocular and monocular views degrades the vivid depth perception on a flat TV. But if we close one eye, Type A is no longer applicable, thus even a flat TV can generate perfect depth perception by Types B-D. For example, one of the authors (KA) temporarily lost vision in the left eye for a few months. After a few weeks of monocular vision, a flat TV screen started to show a truly deep 3D image, especially for video recordings by a drone!

Today's VR headset appears to carry a similar problem. A VR headset includes two tinny LED displays: one for the left eye and the other for the right eye. Two slightly shifted images are projected onto the two displays under the assumption of fixed gaze vectors to a certain depth, but again such an assumption is too naïve. Our vision is the outcome of **MePMoS**. Proper sensorimotor integration between the eyes' motion and visual stimulation is essential to reconstruct 3D images. Without real-time eye tracking and its immediate feedback to the image on the LED screens, true 3D sensation by VR headsets would be severely compromised.

4.10 Evolutionary Aspects: Insects, Rodents, Birds, and Primates

Through this section, we have developed the unified treatment of depth perception by applying **MePMoS**, **NHT**, and **HAL**. Since these are universal principles applicable to any animal, now is a good time to review the evolutionary aspects of depth perception from the origin of vision, starting from insects and other arthropods. We start with arthropod vision as a simpler evolutionary step despite a likely monophyletic origin of all animal eyes (Gehring 2014); significantly more is known about arthropod vision than any other non-chordate phyla. In our view, depth perception is not a luxury of the visual system, but rather has been an integrated part of the vision from the beginning. Once we agree that vision is based on top-down predictions by **MePMoS** like a 3D projection mapping, and agree that vision is expressed holographically by brainwaves for all dimensions by **NHT** and **HAL**, 3D and 2D are essentially the same in complexity.

Let us consider the vision system of insects first. As discussed, recent research showed that insects can observe motion parallax and optical flow to maintain allocentric direction while flying, at least within a 2D flat plane (Apitz & Salecker, 2014; Caves et al., 2018; Kolodkin & Hiesinger, 2017; Nériec & Desplan, 2016; Otsuna et al., 2014; Sato et al., 2013; Mauss et al. 2015). From this observation, it is reasonable to assume that **Type B** – motion parallax and optical follow – is the most primitive and ancient bottom-up mechanism of depth perception. The next step must have been **Type A** – stereopsis of triangulation and binocular disparity, where eye motions are required.

The third step was likely **Type C** – scaling and linear perspective. The essence of Type C (with no motion) is to take full advantage of the scale-invariant Log-polar coordinate system, which only the human has. All other animals have Linear coordinates of [Yaw, Pitch]. It strongly indicates that Type C depth perception is applicable only for humans. Type C has an extreme sensitivity to long-distance, and does not require any motion of the body or eyes. In other words, Type C gives the best predictive power for the largest space-time in the environment, which may have contributed to human's success as hunter-gatherers. **Type D** requires an even higher level of the cognitive inference to judge the consistency between the predicted 3D location/shape and 2D observation. Thus, it must have been the last stage of depth perception, again, specific to humans.

We will revisit and systematically evaluate the evolution of vision systems later in **Section 8.3**.

5 Visual Perception of 3D Body-centric Space

5.1 Overview of the Dorsal "Where" Pathway

This "**Part III:** Holographic Visual Perception of 3D Space and Shape" is focused on answering the fundamental mysteries of human vision below:

1) How we recognize different sizes and orientations of shapes (like human faces)**.**
2) How we faithfully perceive 3D space with depth.
3) How we maintain stable visual perception regardless of saccades.

We have already addressed the first two in **Sections 3** and **4**, which are related to the ventral pathway with scale-invariant log-polar coordinate systems. In this **Section 5**, we will answer the third mystery by exploring the dorsal "where" pathway.

Essentially, our visual perception (i.e., conscious awareness) is three-dimensional and Cartesian-like, which can faithfully represent the external world as it is. Perceived 3D space is not located in our brain but projected into real external space. Even with constant saccadic eye movements, our visual perception is stable and unaltered. Somehow, all our eye and head motions are compensated for to maintain and perceive a stable external 3D space, with our body at the center. This "body-centric" 3D space is the sensation (= awareness) of our conscious vision. How is it possible? The solution must be hidden in the dorsal "where" pathway. The goal of this **Section 5** is to define and derive the body-centric Linear polar coordinate systems in terms of [Yaw, Pitch, Distance].

As already described, in the case of the ventral pathway, 2D retinotopy is known to disappear and be converted to holographic tomography at the stage from V4 → PIT. In contrast, the dorsal pathway keeps 2D retinotopy from MT → VIP/MST → LIP/MST, then all the way up to FEF (Gilbert & Li, 2013; DiCarlo, Zoccolan, & Rust, 2012; Kruger et al., 2013). Especially, LEP and FEF are known to exhibit 2D retinotopy (Schall 1995; Patel et al. 2010; Morris et al. 2012; Morris, Bremmer, and Krekelberg 2013), which sends out the next location for voluntary overt attention to SC (Pierrot-Deseilligny et al. 1995; Engbert 2006).

It is important to note again that, 2D retinotopy by itself is "invisible". Therefore, the observed large Receptive Field (RF) at the Frontal Eye Field (FEF) cannot produce visual awareness yet. Unless the 2D image at FEF is converted to the holographic expression by **NHT**, we cannot perceive any visual image consciously. How and where can we perceive in 3D? It must be at the very end of the dorsal pathway, 7a, where 2D retinotopy from FEF is converted to holographic expression. Thanks to holographic tomography, 3D space can be presented at 7a by the phases of three alpha waves.

Let us review the established static connection of the dorsal pathway in **Figure 6-A**. Once we consider causality and locality, this diagram needs to extend to the time axis and become the space-time "Feynman" diagram in **Figure 6-B** (= a part of **Figure 1-B**). Please note that both figures include the additional connections for overt attention from LIP/FEF → SC, followed by a saccade and corresponding corollary discharge: SC → PN → everywhere in ventral/dorsal visual pathways, which conduct the feed-out of the brainwave for the frame conversion. The most detailed functional space-time connectome is given in **Figures 8-A** (for bottom-up) and **B** (for top-down). As shown here, the dorsal pathway is mutually coupled with the ventral pathway at several stages: PIT ←→ LIP/MIP, AIT ←→ 7a, VTC ←→FEF (Gilbert & Li, 2013; DiCarlo, Zoccolan, & Rust, 2012; Kruger et al., 2013). Thus, these two pathways are not independent but work coherently to form the holographic perception of 3D space and shapes. 7a is the endpoint of their mutual communication, where the complete holographic expression of body-centric 3D space is complete.

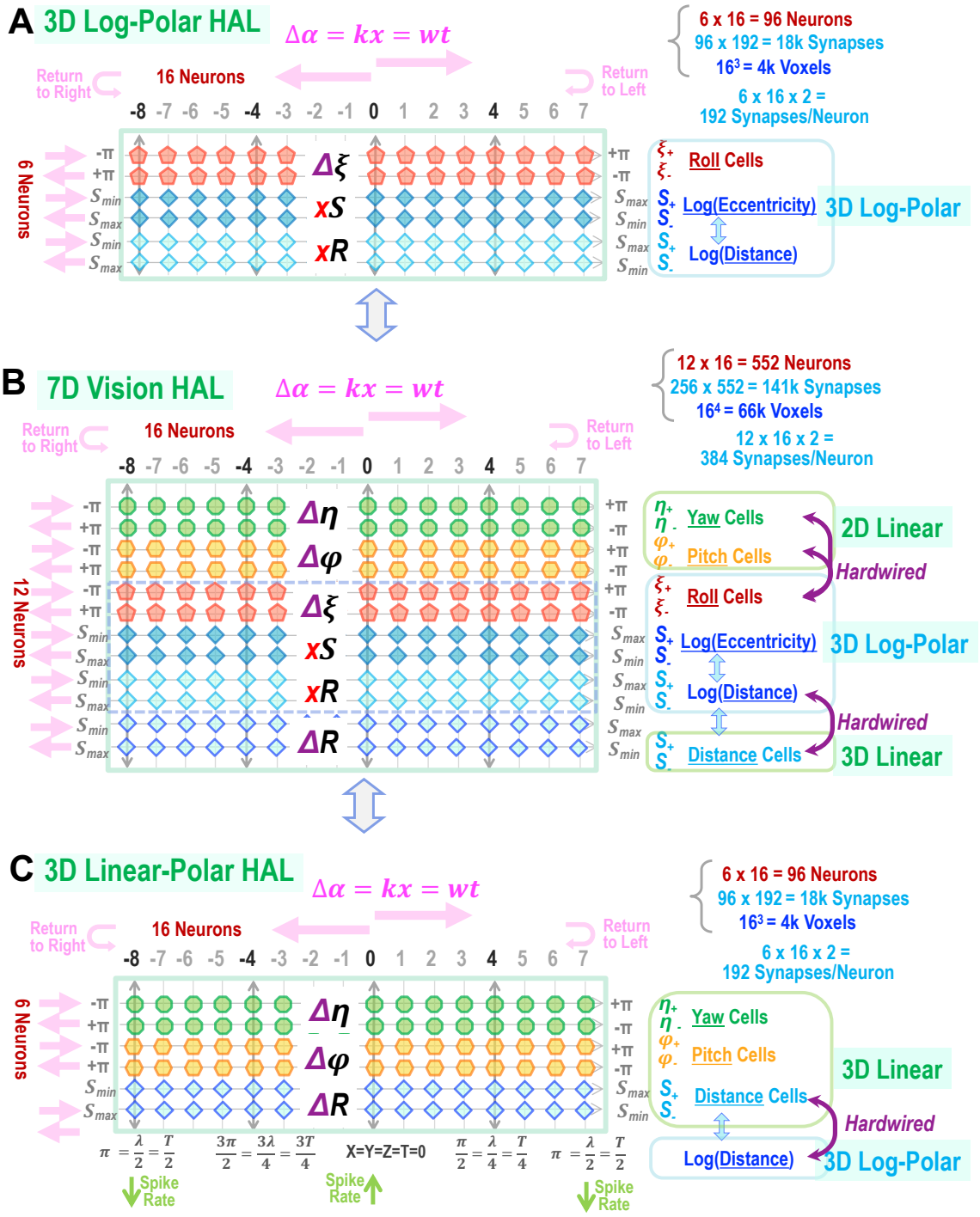

Figure 19. Three types of **HAL** in the Visual Pathways for 7D frame translations. (**A) 3D Log-Polar HAL** from the ventral pathway for scaling and roll. **(B) 7D Vision HAL** at 7a that combines (A) and (B) as one large HAL for 7D frame translation. **(C) 3D Linear HAL** at the end of the dorsal pathway (7a) for conscious visual perception of the allocentric 3D nearly-Cartesian space.

Finally, at 7a, **3D Vision HAL** is formed where we see external 3D space conspicuously. It is like a 3D projection mapping to external space, not by laser scanning beams but by traveling alpha brainwaves.

5.2 Conversion from Egocentric to Body-centric Coordinate System

The end product at 7a is the **3D Vision HAL**. The structure was already presented in **Figures 4** and **5** in **Section 1.5**. Here we will examine the detailed holographic mechanism for frame translation toward the body-centric Linear-polar coordinate system. Our conscious vision is like a 3D projection mapping, which sends out brainwaves (= laser beams) from the center of mass of the head. In other words, our visual perception is free from saccadic eye movements and hard tilting by [Yaw Pitch]. We shall name it "body-centric". Since it is from the body-center radiality going outward, it must be in the Linear-polar coordinate system.

Therefore, 3D Vision HAL must be body-centric Log-polar. That is, both the horizontal and vertical axes are based on eccentricity, representing [Yaw, Pitch] = [Azimuth, Altitude]. On the other hand, depth must be the linear distance from the center of attention with a proper unit of length. And the holographic expression of depth is essential for our visual perception; because of this time expression of depth, the perceived visual image is mapped onto the actual locations of external objects in external 3D space under [Yaw, Pitch, Distance].

Figure 19 shows how the egocentric Log-Polar coordinate system (from the Retina → V1-V4 → Ventral pathway → VTC) will meet with the Linear-polar system from the Dorsal pathway, and formulate linear 3D vision expressed by **3D Vision HAL** at 7a. From the top, **Figure 19-A** shows the 3D Log-polar HAL given by [Roll, Log(Eccentricity), Log(Distance)] along the Ventral pathway. From the bottom**, Figure 19-C** is the Cartesian-like 3D Linear-polar HAL given by [Yaw, Pitch, Distance]. In between, **Figure 19-B** shows the combination of **A** and **C**, which is named **7D Vision HAL**. We believe this 7D HAL is the essence of the frame translation to form body-centric Linear-polar vision.

As shown here, this **7D Vision HAL** consists of six strings of neurons and is capable of 7D frame translations for 3D shape recognition:

1) Yaw (= Azimuth) for horizontal linear translation (and for yaw rotation from body-center)
2) Pitch (= Altitude) for vertical linear translation (and for pitch rotation from body-center)
3) Roll for roll rotation
4) Log(Eccentricity) for scaling up/down
5) Log(Distance) for the direct link to Log(Eccentricity) and to Distance
6) Distance for linear translation into the depth direction

The conversion between distance and Log(distance) is conducted by the fixed hard-wired synaptic network. Assuming a round object of radius R, Eccentricity = R/(Distance). Thus, Log(Eccentricity) = Log(R) – Log(Distance). Therefore the 1D linear transformation between Log(Eccentricity) and – Log(Distance) can be easily achieved by the dual-ring structure of the ring attractor. Conversion between linear distance and Log(Distance) must be hardwired, as is the case of hardwired translation of the retina (= linear distance) to LGN → V1 (= Log(Distance)) in 2D. We assume the same kind of hardwiring for 1D is straightforward.

Next, the conversion from Log-polar to Linear-polar, [Roll, Log(Eccentricity)] → [Roll, Eccentricity], must take place at the entrance of the dorsal pathway, MT, then it stays as it is through MT → MT → VIP/MST → LIP/MST → FEF. Finally, at FEF → 7a, 2D Linear-polar [Roll, Eccentricity] is converted to the Cartesian-like 2D Linear Vision of [Yaw, Pitch]. At this stage, depth is also encoded holographically, forming the final product of 3D Vision HAL of [Yaw, Pitch Distance].

The bottom line is that the **7D Vision HAL** in **Figure 19-B** is intrinsically designed to conduct all kinds of 7D frame translations listed above. Without this holographic expression of 3D visual space by **NHT**, it is impossible to generate and maintain stable body-centric vision in 3D. In a nutshell, the original egocentric 2D retinotopy of [Yaw, Pitch] is holographically perceived at 7a by the body-centric 3D of [Yaw, Pitch Distance]. In the following sub-sections, we will examine the details of covert attention and overt attention, the most fundamental mechanism of scanning external space, while maintaining the stable body-centric frame.

5.3 Covert Attention

How can we identify an important landmark which appears unexpectedly at any location, at any size (due to the different distance), and at any time? This ability is based on attention. Covert attention does not require eye motion, whereas overt attention requires saccadic eye movements. In our daily life, we find an interesting object covertly first, followed by overt attention. Indeed, our life is a series of covert → overt attention steps a few times per second, which we do not notice at all. By doing so, we extract semantic 3D shape information, location by location, while maintaining a stable 3D allocentric (body-centric) frame. How can we perform this so effortlessly? This is the third mystery of vision.

Reading sentences like this paragraph is a perfect example. While we are reading, our eyes must follow the sentence, word by word, by unconsciously moving the eyes from left to right. But if we are forced to stop saccadic eye movement at one spot, say, on top of the word "spot" at the center of the red circle, we cannot read any alphabetic characters beyond the red circle. Our study of peripheral visual acuity showed that we could only read ~10 characters covertly under such a crowded condition regardless of character size (Suri et al. 2020).

We already covered the holographic principle of covert and overt attention in **Part II: Section 3.3**, then briefly reviewed the essential point in **Section 1.4**. Here, we will re-examine the coordination of covert attention and overt attention, illustrated in **Figure 20**. Let us start with the case of a banana, which we memorized once as shown in **Figure 2**, represented by the seven points in 3D: A, B, C … G. The center of the banana was D (X, Y, Z) = (Yaw, Pitch, Distance) = (3, 3, 3). Now we assume a banana has suddenly appeared at D' (X', Y', Z') = (5, 3, 3), that is X' = X + 2 as shown in **Figure 3**.

Covert attention is an internal attempt to shift the banana center from D' to D by X = X'- 2 so that the observed shape can be directly compared with the memorized shape. This is achieved by linear frame translation from the egocentric to the object-centric frame. As shown in **Figure 3**, this example of X' = X + 2 can be corrected by shifting the ring attractor of the X-axis (= Yaw = Azimuth) by 2 units. In **Figure 3**, 3**D Vision HAL** on the upper side is before covert attention, whereas the lower one is after covert attention. If a banana appears at a higher or lower location, similar covert attention along with the Pitch (= altitude) direction will be conducted. By combining these, a banana at any 2D location in [Yaw, Pitch] can be recognized by covert attention.

The horizontal shift of the X (=Yaw) neurons, X = X'-2 in this case, is conducted by shifting the phase of the alpha brainwave by the PN network, described in **Section 2.2**. Besides the regular transfer of the retinotopic image from the retina → LGN → V1, a bypass shortcut exists from the retina → PN. Then PN will generate a vector to the direction of the point of covert attention. This vector is expressed by the phase shift of the alpha brainwave so that the point of interest will become the center of the object-centric frame. This distribution of the alpha phase shift could go even down to V1 as a top-down prediction if succeeding overt attention is anticipated, which ensures the smooth connection of visual perception before/after the saccade.

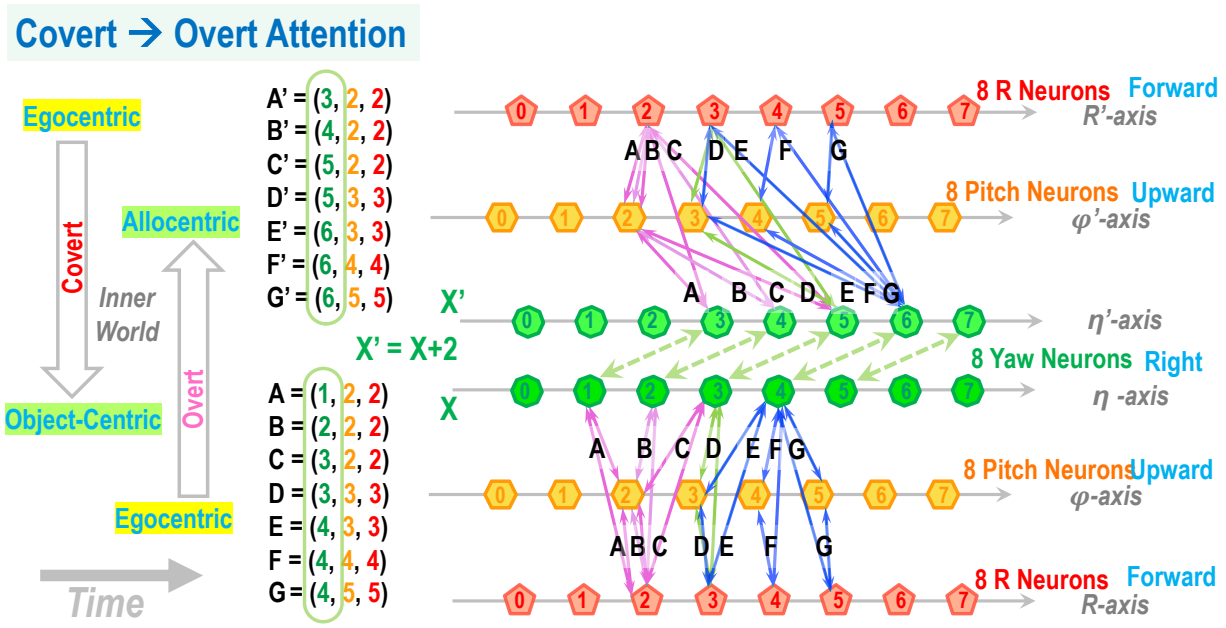

Figure 20. The interplay of covert attention and overt attention. This figure follows **Section 1.3: 1.4** and **Figure 2-3**. The banana shape is originally memorized as Figure 2 with the center D = (3, 3, 3). Then we see the banana the second time, it is shifted to the right by X' = X+2, the new center D = (5, 3, 3). Covert attention transfers the image to the object center by alpha brainwaves and reproduces the original memoried shape. In contrast, overt attention shifts the egocentric image back to the center D = (3, 3, 3) but maintains the off-centered allocentric image by the collolarlay discharge. As a result, our visual perception in our inner world always maintains the stable allocentric frame while extracting the semantic shape at its object-centric frame.

Figure 21-A shows the flowchart of covert attention. Visual stimulation on the retina is directly transferred to PN, then PN distribute the proper phases of the alpha brainwaves to the ventral pathway to reallocate the peripheral landmark to the object center. Then the observed landmark can be directly compared with the memorized one on top of each other by time coincidence.

One should note that this covert attention must be conducted on the Cartesian-like Linear coordinate of [Yaw, Pitch]. Quite intriguingly, even though it is the natural initial coordinate system on the retina for any animal with vision, only the human primary visual cortex converts the retinotopic image to a peculiar Log-polar coordinate system of [Roll, Log(Eccentricity)]. To extract and recognize a shape at a peripheral visual field, our visual pathways must convert [Roll, Log(Eccentricity)] to [Yaw, Pitch].

Why is it the case only for humans? The answer is undoubtedly the scale/rotation invariance of the log-polar coordinate, as explained in **Section 3**. Human vision is optimized to extract semantic shape with any size and rotation, as far as it is perfectly centered on the fovea. On the other hand, a peripheral image is so distorted that it is hard to recognize. In contrast, the vision of any other animal has opposite pros and cons; they can easily identify and recognize the same-size shapes at any peripheral location because the linear translation by covert attention is trivial for their visual pathway. But if the size of the shape is different from the memorized size and shape, even at the central image, scaling up/down to compare with memory must be challenging.

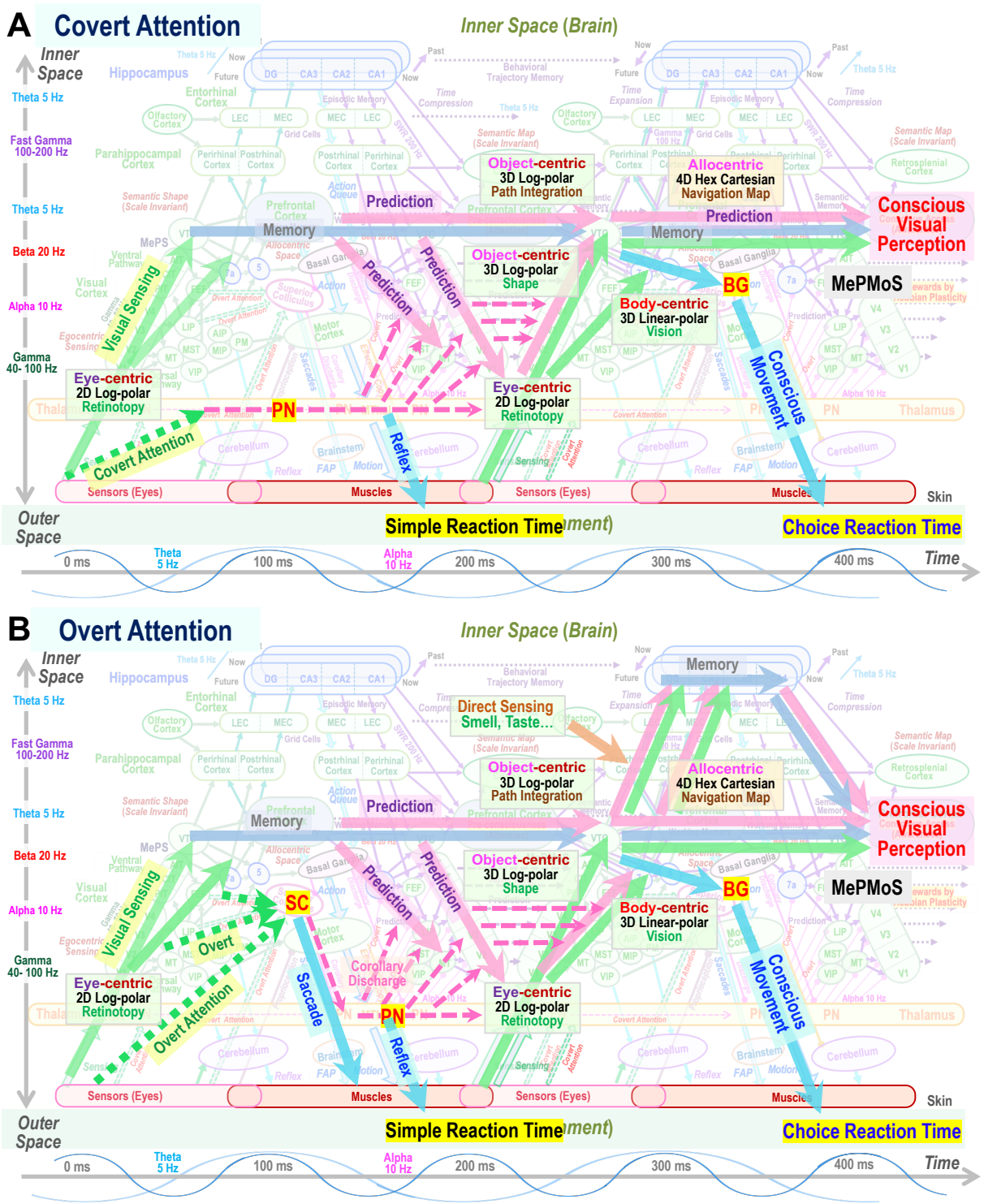

Figure 21. Space-time diagram (flowchart) of covert and overt attention shown in **Figure 1-B**. **(A)** Covert attention is initiated by the bypass circuit from the retina to PN. Then PN distributes the new alpha phases to the ventral pathway to shift the peripheral landmark to the object-center frame. **(B)** Overt attention is triggered by the SC that received the bypass signal from the retina. While SC initiates a saccadic eye movement, SC sends the corollary discharge to PN. Then PN distributes the new alpha phases to the dorsal pathway to maintain the body-centric frame by compensating the eye movement.

5.4 Overt Attention by Saccades – Realization of MePMoS

Now let us consider overt attention. Let us begin with the case of a banana again. Assume that we memorized it some time ago as shown in **Figure 2**, represented by the seven points: A, B, C … G. The center of the banana was D (X, Y, Z) = (Yaw, Pitch, Distance) = (3, 3, 3). Now, a banana has appeared at D' (X', Y', Z') = (5, 3, 3), shifted by 2 units: X' = X + 2 as shown in **Figure 3**. Overt attention re-allocates the center of the banana D' = (5, 3, 3) to the foveal center D = (3, 3, 3). Consequently, we observe the banana at the foveal center in the egocentric frame, the essential process is to recognize the banana shape by overlapping the observed form onto the memorized banana shape.

But we do not notice any image shift in our conscious visual perception. This is due to the corollary discharge – the faithful copy of saccadic eye movement – which compensates for the image shift in real-time. This is coordinated by the PN network by shifting the alpha phase from X (egocentric) → X' = X+2 (allocentric). It is the opposite process from covert attention. In **Figure 20**, overt attention goes up from the lower side to the upper side of the figure. This process allows the extraction of semantic shape of the banana in the newly centered egocentric frame while maintaining unchanged the conscious perception of the 3D allocentric frame. In the end, we continue to observe the banana at the same allocentric location even after eye movements.

The PN network for overt attention is illustrated in **Figure 6**, and the exact process was given in **Section 2.2**. The flowchart is given in **Figure 21-B**. Basically, three independent pathways exist.

1) Fastest: Retina → SC → PN
2) Involuntary: V1-V3 → MT → MST → LIP → PN
3) Voluntary: V1-V3 → MT → MST → LIP → FEF → PN

In the first case, a saccadic eye movement by overt attention is generated directly by the bypass circuit of Retina → Superior Colliculus (SC) → Brainstem (Engbert 2006; Martinez-Conde et al. 2013). Once a new saccade is initiated from SC to the brainstem, a proper corollary discharge is generated and propagated via SC → PN.

The second case is based on the processed image at LIP. If an interesting object is identified at the periphery, then the saccade vector to that location is given to SC involuntarily (and the rest is the same as the first case.) The third case is initiated from the retinotopic image at FEF voluntary. It is reasonable to assume that, in the second and third cases, covert attention is also initiated in parallel so that the target object is shifted to the center of the Log-polar ventral pathways: PIT → CIT →AIT → VTC. (Please refer to **Figure 8**.)

Overt attention precisely follows the **MePMoS** (**Me**mory → **P**rediction → **Mo**tion → Sensing) in this order. The top-down signals of **Me**mory → **P**rediction, expressed by the alpha phase shifts, go down to both the dorsal and ventral pathways. The dorsal pathway converts the memorized and perceived allocentric (body-centered) 3D frame back to the 2D egocentric retinotopy through the top-down pathway: PFC → 7a → FEF → LIP/MIP → MST/VIP → MT → V3 → V2 → V1. Likewise, the ventral pathway propagates the predicted semantic shape at the foveal center down to the ventral pathway in the reverse order: VTC (Fusiform) → AIT → CIT → PIT → V4 → (V3/V2) → V1.

Through these entire processes in both dorsal and ventral pathways, the top-down and bottom-up signals will handshake with each other under the proper coordinate systems at the specific location. This handshake is a perfect example of **MePMoS**, which establishes the mutual communication and coordination to execute the 7D frame translation between the perception at PFC and the sensing at the primary visual cortex V1.

In our daily life, covert attention is usually followed by overt attention. Both are working together synchronously in series. Through this two-step process, allocentric space is maintained, while interesting new landmarks are identified, and their semantic shapes are recognized. As an example, let us try to focus on the apple above first, then focus on the banana next, and go back and forth between them again and again.

As expected, the perceived images of the apple and the banana do not move, as they are attached to the allocentric (body-centric) frame of [Yaw, Pitch] on this paper. No matter how we move our eyes either involuntary or voluntary, we obtain a strong sensation of the semantic shapes + colors of the two objects at their stationary locations. To our best knowledge, **MePMoS** (together with **NHT** and **HAL**) is the first complete theory that describes our stable visual perception with covert/overt attention.

It is worth noting that humans probably require overt attention more frequently than any other animal. Due to the Log-polar coordinate system of our primary visual codex, using covert attention alone it is extremely difficult to identify a new landmark at the periphery. But once it is overtly attended, the landmark is centered on the fovea, and scaling/rotation invariance of the Log-polar V1-V3 helps to recognize the shape tremendously. Clearly, we cannot stop eye motions for the rest of our life. [*Saccadic eye movement is so essential that even Amyotrophic Lateral Sclerosis (ALS) patients utilize it for communication.*]

In this regard, it is worth pointing out that, during REM (Rapid Eye Movement) sleep, we also cannot stop saccades. To generate a visual scene in our dream, we must map it out in front of us by scanning our eyes rapidly, like a 3D projector mapping. After all, dreams and real vision share the identical principle.

Lastly, it is well-known that alpha brainwaves become prominent when we close our eyes, but they fade away once we open the eyes. It is probably related to the constant alpha phase shifts during covert and overt attention while eyes are opening. But once eyes are closed, no attention is needed; thus, the alpha brainwaves become stable and synchronized.

5.5 Attending to and Pursuing Moving Objects

Our vision is remarkably sensitive to moving objects. Any slight motion of a tiny object, like a flying insect, triggers our covert attention even in the far periphery. Consequently, we tend to focus on it unconsciously by overt attention, and then start to chase its movement effortlessly. During our pursuit, the moving object continues to appear stationary on the center of the fovea, while the background view is flowing in the opposite direction in the egocentric frame on the retina. But what we visually perceive is the actual motion of the object in front of the stationary background within the allocentric frame. Such a perception of moving objects in the allocentric frame must have been essential for any animal to chase prey and escape from predators, as prey and predators are moving around.

Clearly, the conversion from egocentric 2D retinotopy to allocentric 3D visual perception must occur all the time, which we have discussed extensively already. The key is the concept of **MePMoS** and **NHT**; Our vision is the internal holographic projection of the predicated 3D image, like a 3D projection mapping. This 3D mapping is within the allocentric, Cartesian-like, Linear-polar coordinate system after compensating for eye/head motion.

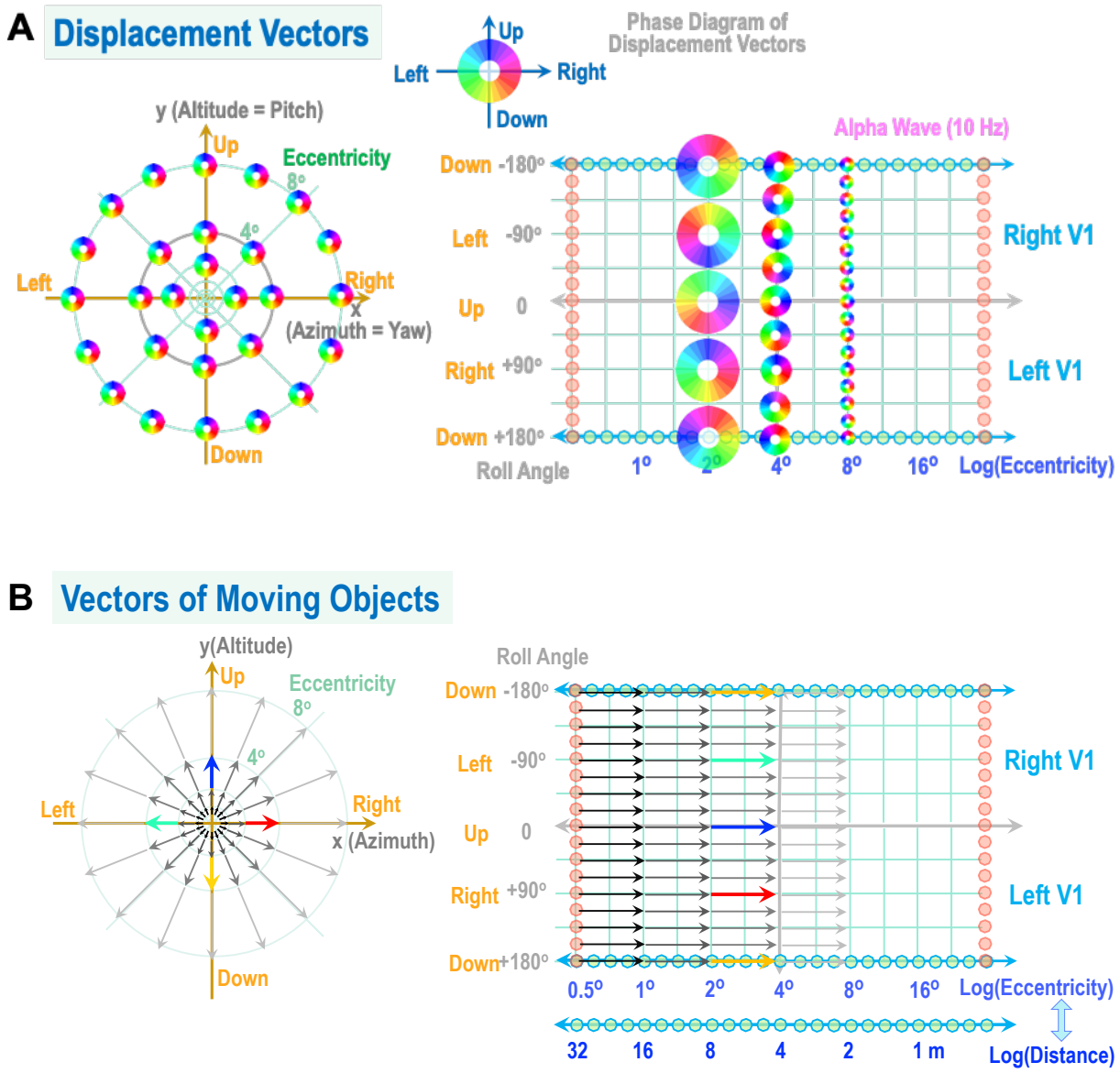

Figure 22. Mapping of moving objects from the retina to the Log-polar SC and V1. **(A)** is the phase diagram of the displacement unit vectors. This shows the basic mapping of the unit vector at any local 2D location in [Yaw, Pitch] (left) onto the Log-polar coordinate of SC and V1(right). **(B)** shows the vectors caused by radially moving objects, which are displaced by a factor of two in eccentricity. Once these are mapped to the Log-polar V1, horizontal patterns are detected by the orientation column. The foveal center is extremely sensitive to displacements as small as one degree or even less.

The basic mechanism of the detection and pursuit of moving objects is similar to depth perception by moving the eyes (Type A) or moving the body (Type B), discussed in **Section 4**. Depth perception is based on the host animal's own motion to view a stationary environment. In contrast, in the case of pursuing moving objects, a host animal is at rest and the target is moving.

Let us consider the pursuit of moving objects in detail, following the three steps below:

1) Step 1: Covert attention to a moving object at the periphery.
2) Step 2: Overt attention to the object for relocating it to the foveal center.
3) Step 3: Pursuing its motion while keeping it at the foveal center.

Step 1 is covert attention to a moving object at the periphery. **Figure 22-A** shows the displacement vector caused by a moving object, assuming the same apparent speed on the 2D visual space at the retina: [Yaw, Pitch]. Like **Figure 15-A**, the direction is given by the color phase diagram. Near the foveal center, the sensitivity to a time movement is dramatically enlarged on the Log-polar V1. At **Step 2**, overt attention is activated (by Retina → PN, LIP → PN, or FEF → PN), and the target object comes to the foveal center. Finally, at **Step 3**, the eyes are locked on the target and begin to pursue the moving target.

Once our eyes are locked on the target, then the target always moves radially outward from the foveal center. This radial displacement vector from the foveal center is mapped onto the Log-polar SC and V1, expressing the enlarged horizontal displacement vector as shown in **Figure 22-B**. At the foveal center, a tiny displace vector is dramatically enlarged on V1 as a horizontal vector. The orientation column in V1 is optimally designed to pick up such horizontal vectors. This tells us the superiority of the Log-polar V1 that is specific to humans.

This horizontal vector to the right on V1 is extremely clean against the quiet background scene. Consequently**, Step 3** of the pursuit is initiated promptly. The displacement vector, representing the pursuit of the target, seems hardwired to the SC for automatic saccadic movement. The critical consequence is that SC not only moves the eyes but also sends the corollary discharge to compensate for eye motion in the egocentric frame to recover and maintain the allocentric frame for visual perception.

As a result, we continue to maintain the stationary background under the allocentric frame (as is the case of any overt attention.) At the same time, the moving target at the foveal center is perceived with the proper motion in the allocentric frame. Basically, our saccadic eye movements automatically give the trajectory of the target in our 3D vision. That is exactly the outcome of **MePMoS**. Our vision is predicting the trajectory of the target by saccades, which is the visual sensation by the projection mapping.

5.6 Summary – Maintenance of the 3D Body-centric Frame

In this **Section 5**, we have discussed how 2D retinotopy under the egocentric frame of [Yaw, Pitch] can be converted to 3D visual perception under the body-centric frame of [Yaw, Pitch, Distance]. The challenge is that the human's primary visual cortex V1-V3 is under the Log-polar coordinate system of [Roll, Log(Eccentricity)]. Therefore, two steps of conversions are necessary:

1) Ventral: [Yaw, Pitch] → [Roll, Log(Eccentricity)] → [Yaw, Pitch, Roll, Log(Distance)]
2) Dorsal: [Yaw, Pitch] → [Roll, Eccentricity] → [Yaw, Pitch, Roll, Distance]

In our model, the ventral and dorsal pathways are integrated for mutual commutation at region 7a, where the fourth axis can be expressed and hardwired by the four types: Distance, Log(Distance), Eccentricity, and Log(Eccentricity), as shown in **Figure 19**.

This arrangement is optimized for the prompt 7D frame translations of

1) 1D scaling,
2) 3D rotation of [Yaw, Pitch, Roll], and
3) 3D linear translation of [Yaw, Pitch, Distance], and

to achieve two goals:

1) Extraction and recognition of the 3D semantic shape of landmarks, and
2) Maintenance of allocentric 3D space.

We specifically discussed three cases that occur in this order:

1) Covert attention
2) Overt attention
3) Pursuit of a moving object, while maintaining it at the foveal center

Thanks to the unconscious, spontaneous frame translation from egocentric 2D vision to the body-centric 3D frame, our eye and head motion is compensated for and unnoticed. But once we start to move our body (to chase prey), then to be exact, the body-centric coordinates cannot represent the truly allocentric Cartesian coordinates. This is the last step of conscious visual perception of external 3D space for navigation. We will discuss this in the following **Section 6**.

6 From Body-centric to Allocentric, Linear-Polar to Cartesian

6.1 From Body-centric to Allocentric, from Linear Polar to Cartesian

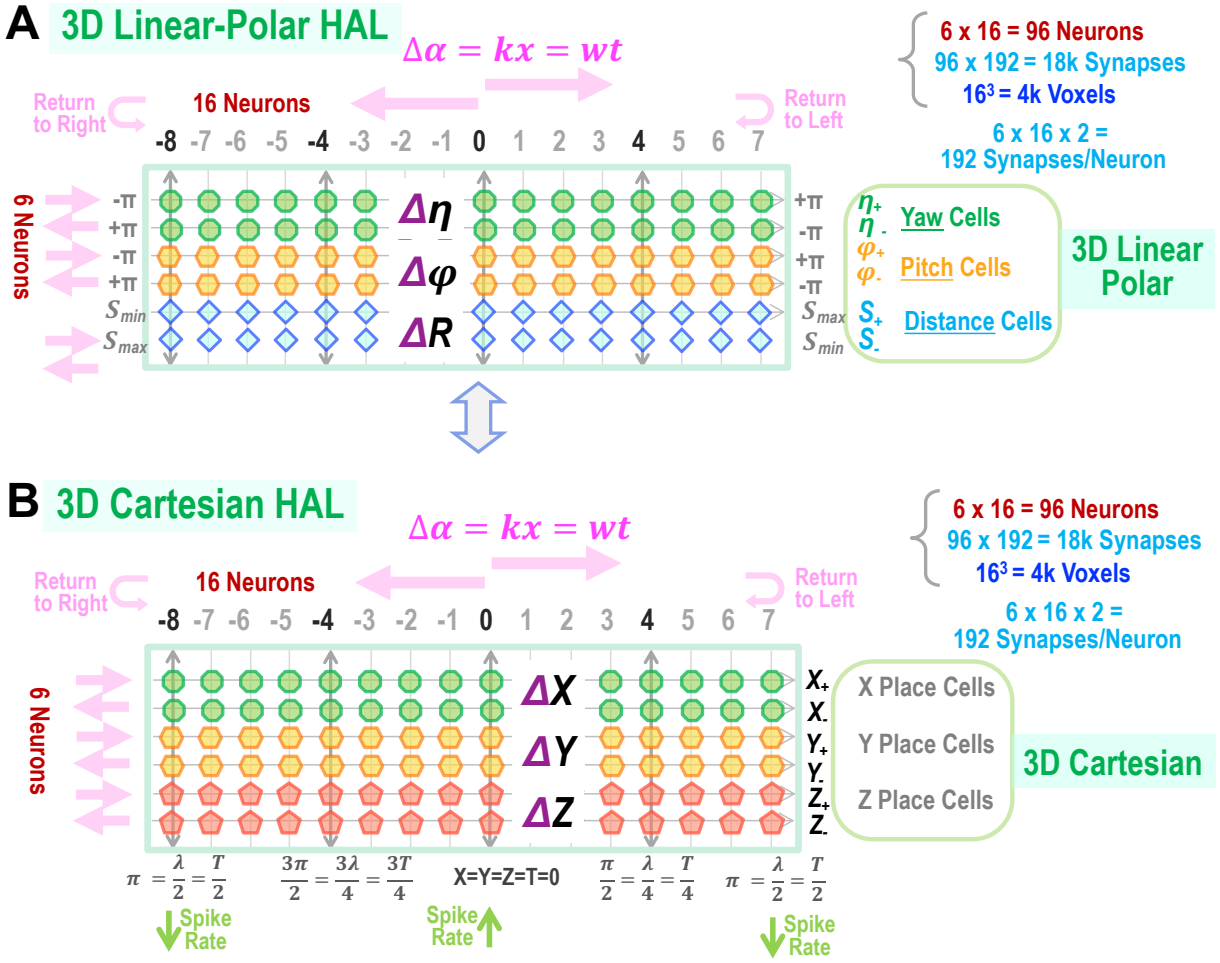

Figure 23. **(A)** Body-centric **3D Linear-polar HAL** expressed by [Yaw, Pitch, Distance]. **(B)** Truly allocentric **3D Cartesian HAL** expressed by [X, Y, Z].

Ultimately, we must navigate given 3D external allocentric space to find and eat good food and come back home. And our navigation in space heavily relies on vision. Even though visual perception is based on the body-centric 3D Linear-polar system of [Yaw, Pitch, Distance], somehow, it must be transformed to the allocentric 3D Cartesian system of [X, Y, Z]. In fact, once we start to move around in space, we seem to sense the Cartesian system; A table is a square, a bedroom is a square, a street has a fixed width, and so on. These are all sensations of Cartesian space.

In this regard, insects' navigation is not an exception. Recently, the biological mechanism of the Polar-to-Cartesian frame transformation has been studied in Drosophila, and it was identified in 2D

navigation (Lyu, Abbott, & Maimon, 2020). (Detail was given **in Part II: Section 1.3**.) Here, we assume that a similar mechanism was inherited by vertebrates including humans.

Figure 23 illustrates such a frame translation from the Linear-polar system to the Cartesian system in 3D. **Figure 23-A** is the body-centric **3D Linear-polar HAL** expressed by [Yaw, Pitch, Distance], which is the outcome of the human visual pathway at 7a. Then it must be transformed to the allocentric **3D Cartesian HAL** shown in **Figure 23-B**, probably following the biological mechanism similar to the 2D translation in Drosophila (Lyu, Abbott, & Maimon, 2020). We assume this happens somewhere at around 7a → Parahippocampal Cortex.

The final step of visual sensation by the Cartesian system is the critical input to the Hippocampal system for navigation, which we will discuss in **Section 6.5** and **Part IV.**

6.2 Body Motion in 3D Allocentric Space

Finally, we are ready to complete the model of vision-based navigation in external 3D space. The essential point is that we manage to sense allocentric 3D space faithfully regardless of our eye, head, and body motions. And this is the direct outcome of **MePMoS** and **NHT**. Since we already covered eye/head motion in the previous section, let us complete the argument by considering body motion in 3D: to the left/right, up/down, and forward/backward directions.

To begin with, **Figure 24-A** is the linear frame translation for the side-body motion. It is again based on a banana that suddenly appears at the center D' = (5, 3, 3). To eat it, we move to the right by two units, then we will observe the banana center at D = (3, 3, 3), but the allocentric frame remained the same, still showing the banana center at D' = (5, 3, 3). It is essentially identical to the covert → overt attention given in **Figure 20**. Instead of saccadic eye movement, we physically move our body in this case. Visually it is true that the center of the banana is shifted to D = (3, 3, 3) in the body-centric frame. But at the same time, our conscious perception is telling us that it is not a banana, but we moved our body. That is the dual conscious sensations: the body-centric 3D vision and the allocentric 3D space.

Next, **Figure 24-B** shows the case of forward body motion by one unit: Z' = Z + 1. In this case, we move forward to come closer to each banana. Initially, the center of the banana is D' = (3, 3, 3), after the forward motion, it will become D = (2, 3, 3). In our vision, we observe a banana showing up closer and larger. But at the same time, our conscious perception is telling us that it is the same banana with the same size, located at the same allocentric place of D = (3, 3, 3). Once again, this is the dual conscious sensation.

In conclusion, no matter how we move our body center in 3D – to the left/right, up/down, and forward/backward – the location of the banana is unchanged and always attached to the allocentric Cartesian coordinate system. The underlining biological mechanism has been inherited from the ring attractor from the common ancestor of chordates and arthropods, like the ring attractor found in extant species such as Drosophila (Lyu, Abbott, & Maimon, 2020). Such stable sensation of the banana at the same location within the allocentric frame is the origin of "**place cells**" in Hippocampus, which is the topic in **Section 6.5** and **Part IV.**

It is also important to reaffirm that any movements – physical body motion mentioned above, overt attention by saccadic eye movements, or head reorientation – share the identical mechanism to maintain allocentric 3D space intact, which is another notable aspect of the **Grand Unification of mind and brain**.

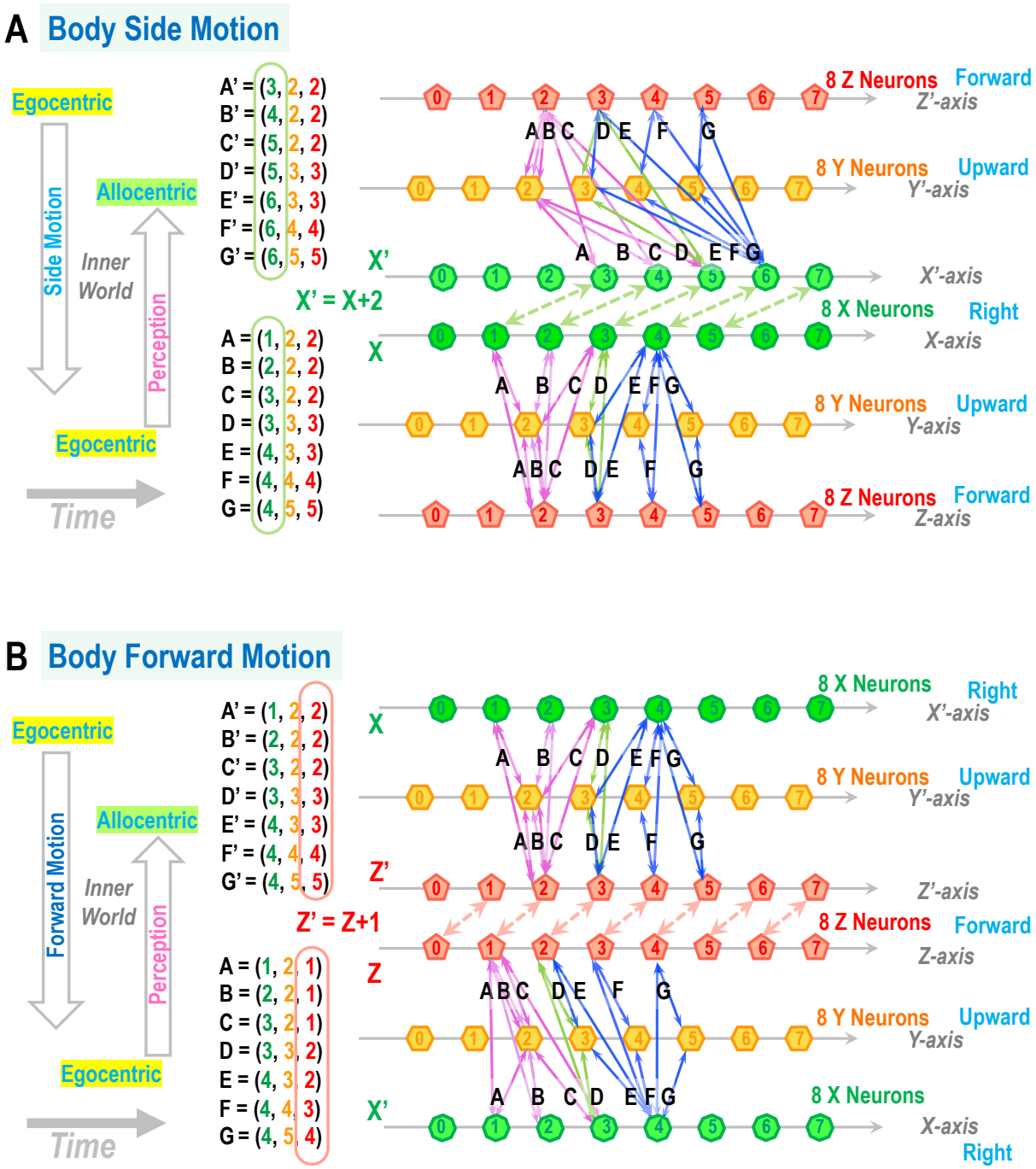

Figure 24. Maintenance of the allocentric 3D Cartesian HAL during body motions. **(A)** is the case when a host animal moves to the right by two units, X' = X + 2, while maintaining the allocentric frame unchanged. **(B)** The case when a host animal moves forward by one unit, Z' = Z +1, while maintaining the allocentric frame unchanged.

6.3 Chasing Moving Objects

Finally, we can discuss the most realistic, complex behavior of animals in daily life. As an example, if a wolf finds a moving rabbit, the wolf rushes in the direction of the escaping rabbit and continues to chase it. During this period, the wolf continues to sense the proper motions of the rabbit within allocentric 3D space, while keeping track of its own location in the allocentric frame, very much like a GPS. Evolutionarily, such execution of chasing prey must have been the strongest pressure to advance the visual system, thus driving brain evolution. From the rabbit's point of view, escaping well in the allocentric frame is equally or even more critical. It's a question of life or death.

One should note that the process of chasing prey generates quite a vivid depth sensation thanks to optical flow and motion parallax, given in **Section 4.6**. And vivid depth perception is quite applicable to locating the moving target in the allocentric 3D environment. We can observe the clear 3D trajectory of escaping prey as a function of time.

However, the above strategy and procedure of chasing moving targets seems to require extensive neural processing power. Many animals – such as insects, birds, and fish – freely navigate open 3D spaces either in air or water. Their strategy of chasing prey could be much simpler. To accomplish the mission, they do not have to reconstruct allocentric space; what is necessary is to chase prey in the egocentric 3D polar coordinate system. At least, this should be the principle of the final attack.

6.4 Optical Flow by Passive Body Motion

This explains why we can drive a car safely even surrounded by all kinds of moving objects, such as in traffic crowded with cars, bicycles, pedestrians, etc... The optical flow will locate our car on the allocentric map (like a GPS), then any other moving objects are also immediately identified on the allocentric map. Playing sports, like football or basketball, is the same. Thanks to the active movement of the players themselves, they can vividly and reliably sense the 3D location of moving balls and other players within allocentric 3D space. When we receive a served tennis ball, we'd better keep moving our body left and right rhythmically to enhance the depth perception of the incoming high-speed ball by motion parallax.

But according to the strict concept of **MePMoS**, optical flow can be treated properly only if it is caused by the host animal's motion by muscle contraction, thanks to the corollary discharge. When we drive a car, the corresponding optical follow is caused not by our body motion but by the car's motion, while we are sitting on a driver's seat (thus no corollary discharge.) Why is it possible to maintain the perception of stable allocentric space and sense our car's forward locomotion like a GPS?

To understand this trick, we do need to consider the origin of corollary discharges caused by muscle contractions for body motion. As shown in the space-time diagram, **Figure 1-B** and **Figure 6**, the Basal Ganglia (BG) gives the final go/no-go commend to the motor cortex. At the same time, the same BG generates a corollary discharge to PN (similar to SC → PN in the case of saccades.). We also assume that initial retinotopic images at LIP will give the opposite of the optical flow vectors (as the host's own motion vector) to PN. Therefore, maintenance of a stable allocentric frame should come from the agreement of the two pathways to PN.

1) BG → PN: Corollary discharge of body motion
2) LIP → PN: Optical flow and motion parallax due to body motion

The coincidence of these pathways at PN will generate the appropriate phase shifts of alpha brainwaves to maintain the visual sensation of allocentric 3D space. When we drive a car, (1) is missing, which causes confusion at PN. To avoid this, we hypothesize that BG is well designed to emulate the car in motion and generate a "virtual" corollary discharge to PN. This could explain why

we do not get motion sick as a driver, but often get it as a passenger. If we are involved in driving a car, we can emulate the car's motion at BG much more easily than as a passenger.

Or when we are seated on a train at rest at a platform, if an adjacent train outside the window starts to move, we often get confused about whether our train or the other one has started to move. With the virtual corollary charge from BG, we can move our train. But without it, the other train would move instead. This means that the successful creation and maintenance of the allocentric frame depends on the proper integration of the corollary discharge by BG → PN and optical flow by LIP → PN.

6.5 7D Frame Translations – Recognizing Multiple Landmarks

We have been emphasizing the necessity of the 7D frame translation to recognize objects at any 3D location, any size, and any 3D orientation. The 7D frame translation seems absolutely necessary to recognize the once-memorized 3D shape that could appear at any 3D location, with any size, and with 3D orientation in [Yaw, Pitch, Roll]. Without such 7D mental translation, a newly observed object cannot be directly mapped exactly on top of the memory to create the coincidence at every single neuron that forms the memory unit, engram. This is the only way to satisfy causality and locality. For this very reason, we have developed the new concept of **MePMoS**, **NHT**, and **HAL**.

Let us reexamine how far we have successfully accomplished the above. Below is the checklist of our achievements.

1) 1D scaling up/down by the scale-invariant Log-polar **HAL: Section 3.2**
2) 1D "Roll" rotation by the rotation-invariant Log-polar **HAL**: **Section 3.2**
3) 2D liner translation in the Cartesian-like 2D space of [Yaw, Pitch] = [Azimuth, Altitude] by covert and overt attention: **Sections 5.4** and **5.4**
4) 1D depth translation by utilizing Log(distance) = - Log(eccentricity): **Section 4.7**

Although this is an impressive list, we still have to complete two missing degrees of freedom: 2D rotations of the 3D object by [Yaw, Pitch]. This is a daunting task because 3D rotation can be performed naturally only under the 3D polar coordinate system with the center of the rotation located at the center of the 3D object. To achieve this, the center of the coordinate system must be linearly transformed in the 3D Cartesian coordinate system from the observer to the target object.

Fortunately, the required neural system for such a 3D linear translation evolved as the insect-like ring attracter and is also emended in the human vision and navigation system as described above. Therefore, we assume that 3D rotation of the object is effectively conducted in three steps:

1) 3D linear translation of the Cartesian coordinate center from the observer to the object.
2) 3D rotation by the 3D polar coordinate in [Yaw, Pitch, Roll]. In particular, 2D rotation by [Yaw, Pitch].
3) 3D linear translation of the Cartesian coordinate center from the object back to the observer.

In the end, the observer can visually perceive the rotated object with any 3D orientation. This is the most complex operation required to recognize 3D objects at any location with any orientation. Since we can reliably recognize numerous human faces and different car types, the above three steps must have been encoded around the cortical region 7a for 3D object recognition. Finally, the mission of the 7D frame translation has been accomplished.

In our life, we are always surrounded by multiple 3D objects. We must promptly recognize many of them at their 3D locations within the allocentric frame. One by one, we pay overt (or covert) attention to each object and process its 2D retinotopic image to create a 3D shape, then conduct the above 7D frame translation to compare the 3D shape with memorized 3D shapes of similar objects.

All these processes are conducted effortlessly and unconsciously. Only the outcome – the recognized semantic information of multiple 3D objects, attached to the perceived stable allocentric 3D linear space – comes out to our conscious awareness. After all, our vision is an astonishing product of nature.

6.6 From 3D Visual Pathways to Hippocampal Networks

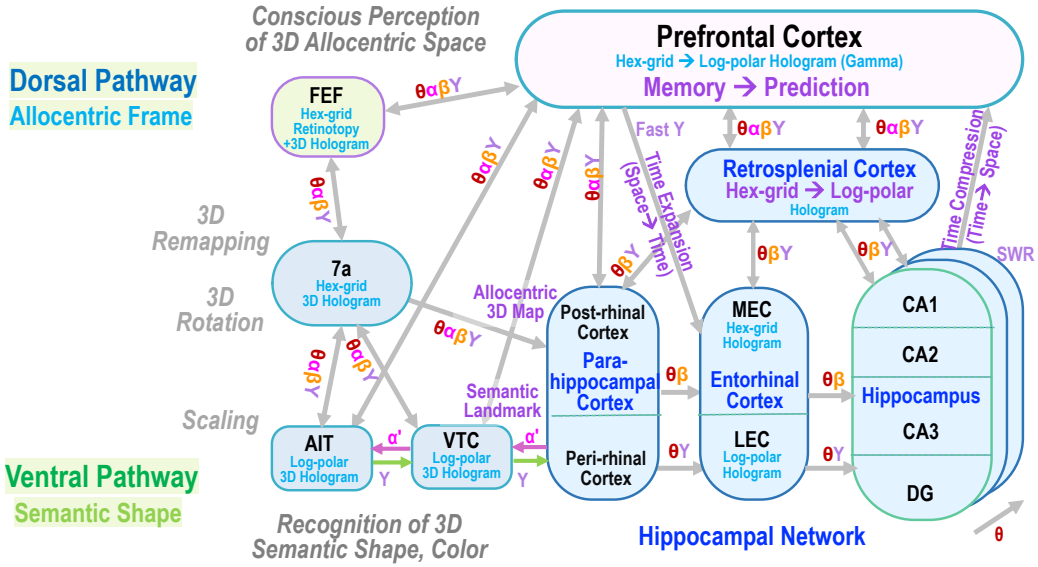

Figure 25. Connections from the visual pathways to the Hippocampal network. The dorsal pathway (FEF → 7a) generates the "predicted" visual perception of 3D space, which is injected into the Postrhinal Cortex → MEC (Medial Entorhinal Cortex) → Hippocampus. The ventral pathway (AIT → VTC) generates the semantic (scale-invariant) 3D shape of a landmark, which is injected into Perirhinal Cortex → LEC (Lateral Entorhinal Cortex) → Hippocampus.

3D vision is not the end goal of our brain. Ultimately the brain is designed to navigate space in the direction of good food (or escape from predators). To achieve this final goal, the outcome of visual processing mentioned above will be injected into the Hippocampal network. The corresponding signal pathways are given in **Figure 25** (which is extracted from **Figure 8-A**.) Fundamentally, there are two distinct inputs into the Parahippocampal Cortex. Firstly, the reconstructed allocentric **3D Cartesian HAL** is injected into Post-rhinal Cortex → MEC, which generates the GPS-like map by the grid cells and place cells. Secondly in parallel, the recognized landmarks with 3D semantic shape information – in form of **3D Log-polar HAL** – are injected into Peri-rhinal Cortex → LEC. For both pathways, **7D Vision HAL** at 7a (in **Figure 19**) is playing a critical role in the holographic information transfer.

Here is one critical mismatch to be considered. All the visual signal processing for 3D perception is conducted by the alpha brainwaves. However, it is well known that Hippocampus is governed by the theta brainwaves (~ 5 Hz). Here again, the beauty of the **HAL** is that it consists of a static lattice structure with geometrical 2D synaptic connections; thus, it is fundamentally frequency independent. In other words, at 7a → Parahippocampal cortex, there must be natural frequency down-conversion from alpha waves (~10 Hz) to theta waves (~5 Hz). These **HAL**s are written by the alpha and read out by theta.

Speaking about multiple frequencies in the brain, to be precise, there is bottom-up local image parallel processing by the gamma band, including color, local shape, and depth (which will be explored in **Part V**.) We already explained the local depth sensation generated by high-frequency gamma in **Section 4.3**. Therefore, depth perception involves three steps:

Step 1) Gamma (~100 Hz): Bottom-up local image parallel processing for depth.
Step 2) Alpha (~10 Hz): Body-centric **3D Linear-Polar Vision HAL** for visual perception.
Step 3) Theta (~5 Hz): Allocentric **3D Cartesian HAL** for navigation

This could explain the origin of our conscious awareness by the dual coordinate systems. On one side, our vision follows the body-centric 3D linear-polar coordinate system. But at the same time, we can sense that the external environment is allocentric and Cartesian. To avoid interference between these two, like today's WIFI technology, our brain assigns two distinct frequency bands for the dual sensations of external 3D space: one (~10 Hz) for vision, the other (~5 Hz) for navigation.

6.7 Summary – Maintenance of 3D Allocentric Cartesian Frame

In this **Section 6**, we have worked through another remarkable aspect of human vision. We seem to construct the dual coordinate systems for our visual conscious awareness in 3D space.

1) Body-centric Linear-polar coordinates of [Yaw, Pitch, Distance] = [Azimuth, Altitude, Distance].
2) Allocentric Cartesian coordinates of [X, Y, Z].

This concept is consistent with our daily life experiences. When we are viewing external space only by moving the eyes and head from the fixed body center, we do not notice any apparent motion of external 3D space. That is body-centric Linear-polar coordinates of [Yaw, Pitch, Distance]. But as soon as we start to move our body (= the center of the head), the apparent vision is going to shift by motion parallax and optical flow. However, interestingly, our conscious mind tends to create the other stable coordinate system for navigation. That is the allocentric Cartesian coordinates of [X, Y, Z].

The second visual perception from the allocentric Cartesian system is required to navigate external space reliably. This is the injection to the navigation system of the Hippocampal network. This second sense of the Cartesian must be coordinated by theta brainwaves. Such dual sensations by the dual-frequency bands seem a key feature of our brain. The following **Part IV** will fully be devoted to the Hippocampal navigation system by theta brainwaves.

7 Experimental Evidence of MePMoS and NHT – Reaction Time

7.1 Prediction through MePMoS and NHT in 3D Visual Perception

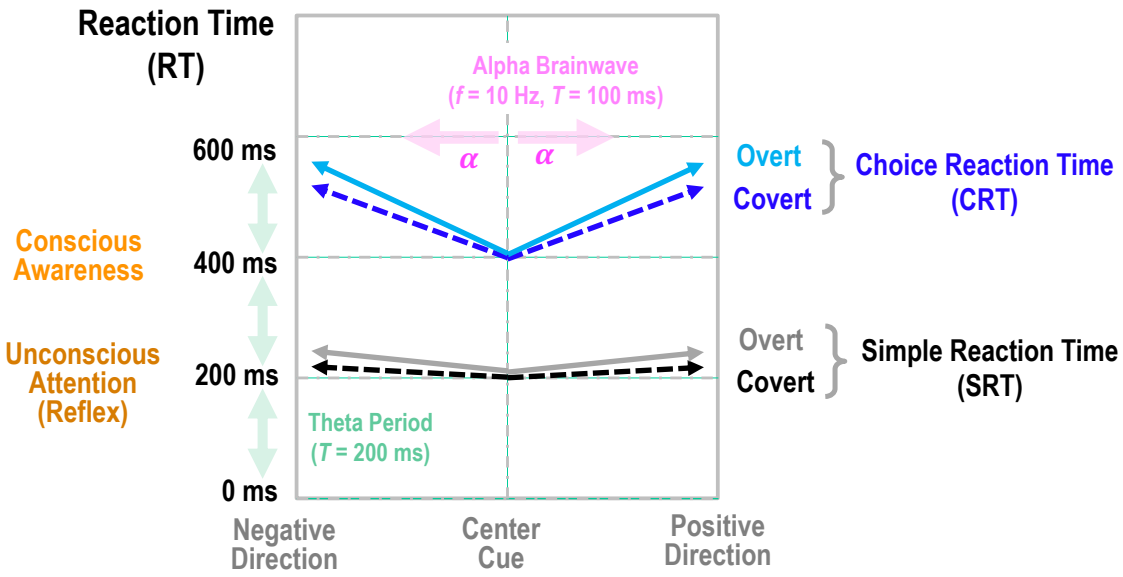

Figure 26. Prediction of Reaction Times (RT) for four conditions: Simple vs. Choice RT and Overt vs. Covert RT. Simple RT involves one type of visual stimulation, and therefore accounts for a reflex completed in one theta period (~200 ms). This process does not create any sensation of space. Choice RT accoounts for the conscious decision-making in face of multiple options, and therefore consists of at least two theta cycles (~400 ms). Overt RT and Covert RT must be nearly identical, as we do not notice saccades in our daily visual perception. For the CRT task only, any deviation of the stimulation from the center cue (i.e., the horizontal axis) leads to an increase in the RT of up to one alpha brainwave period (~100 ms) that is proportional to the magnitude of deviation.

The models we have developed – **MePMoS**, **NHT,** and **HAL** – all prompt one specific conclusion: space is an illusion (invisible, to be exact) and it must be converted to time in order to be perceived. Our theory suggests for the indisputable nature of this notion, which should be able to be explicitly proven by experimental evidence.

Let us take a simple example. If a one-foot-long bar is horizontally placed in front of us, the visual sensation of its length will be derived within a specific time span. Alternatively, if the bar is two-foot-long, it would take twice the amount of time for such perception of the length to arise. This phenomenon is imposed by the constraints of causality and locality conceptualized by Einstein, which is subsequently affirmed by Hebb and could be conveniently verified by a convenient classic reaction time (RT) experiment. To exemplify this, when a subject is asked to visually focus on the center of a big TV screen with coordinates [Yaw, Pitch] = [Azimuth, Altitude] = [0, 0], a spot is to be flashed at a random time and peripheral location (e.g., Yaw = Azimuth = 0, ±10°, ±20°, ±30°…). Then, the RT the subject takes to perceive the flashing light is expected to be delayed proportionally with the eccentricity (i.e., Azimuth, or Yaw) of the speck of light.

We have elaborated on this prediction to design more quantitative experiments, as shown in **Figure 26**. Here, the horizontal axis indicates the deviation (displacement) from the center in any one of the 3D spaces, 1D scale, or 3D rotations within the 7D phase space. Roughly speaking, any deviation in 7D space would introduce an extra RT that is proportional to the magnitude of deviation. However, various conditions should be carefully considered, as they could yield systematic differences. If the stimulation, such as a flashing light, is of only one semantic type and does not prompt any decision making, this information can be processed by a mere reflex, which can be accounted for by Simple Reaction Time (SRT). In this case, we can react unconsciously within around one theta period (~200 ms), regardless of the deviation from the cue. In contrast, if the stimuli shown were selected from a set containing more than one semantic type (e.g., color, shape, pattern, etc.), a conscious decision with respect to semantic content must be made; thus, Choice Reaction Time (CRT) – which is applicable here – requires conscious awareness, which takes at least two theta cycles (~400 ms).

Such quantization by the theta period is the direct outcome of **MePMoS**, which is illustrated in the space-time diagrams in **Figures 1-A** and **B**. In the case of CRT, an additional delay proportional to the magnitude of deviation from the center of vision is expected. This delay is caused by the requirement for the memory to be shifted from the center to the location of the incoming stimulation at the periphery – a process facilitated by shifting the phase of the alpha brainwaves. The required time for this phase shift accounts for the sensation and measurement of the location of the observed object. Finally, the space-to-time conversion is complete. Since this conversion is conducted by the alpha brainwave, we expect the extra delay in CRT equivalent to the period of the alpha of up to 100 ms.

SRT and CRT can both be measured seaparately under either covert attention or overt attention. However, Overt RT and Covert RT should be nearly identical, as saccadic eye movements are unnoticeable throughout our conscious visual perception of 3D space. Provided that our perceived vision is stable even with saccades and that space must be converted to time, then the time to express the space (i.e., RT) must be also stable with saccades. **Figure 26** incorporates all these facts. The Equivalence of Overt RT and Covert RT is an essential prediction, which means that:

1) During overt attention, the eyes are moving at a constant speed, such as ~100 ms for 60 degrees.
2) During covert attention, the alpha brainwave is traveling at the same constant speed.

Designing RT experiments that systematically incorporates all 7D frame translations is uncomplex. For this purpose, experimental designs that study the recognition time of unfamiliar human faces provide the most realistic conditions. In reality, a face of any size can appear at any 3D location, with any 3D orientation. If we only memorize human faces at the fixed size and the fixed distance in front of us, such as one of $\pm 4^{o}$ (i.e., 8^{o}) height at a distance of 1.4 m (as shown in **Figures 11** and **17**), then processing the same face that is presented on a large TV screen – whose size, eccentricity, and orientation deviate from those in the original memory – will require extra reaction time; the length of this delay should be proportional to the magnitude of the deviation from the original memory. See below the complete list of the 7D frame translations:

1) Horizontal eccentricity: Yaw = Azimuth (~X) (in degree)
2) Vertical eccentricity: Pitch = Altitude (~Y) (in degree)
3) Distance (R) ~ Depth (Z) (in meter)
4) Size in Log (Height /8^{o})
5) Rotation in “Roll” angle (in degree)
6) Rotation in “Yaw” angle (in degree)
7) Rotation in “Pitch” angle (in degree)

We can test for all of the above 7D (except distance) by measuring SRT for only one face or CRT for two or three faces (= CRT) in the stimulus set. The designation of either covert or overt attention to the faces can be controlled for through specific instructions for participants of the experiments.

A TV-based experiment is not only easy to conduct, but also extremely flexible in displaying visual stimulation, such as human faces. Unfortunately, the temporal resolution of a large TV is sub-optimal, due to coarse quantization by the frame rate (mostly 60 Hz, resulting in ~17 ms). An alternative strategy utilizes flashing patterns of discrete LED strips or LED arrays; the temporal resolution with this strategy is far superior (< 1 ms), at the cost of limiting the flexibility of the semantic patterns generated. On the other hand, the depth perception can be reliably studied.

7.2 Experimental Results – Reaction Time under 7D Translations

Considering all the above assertions, we have conducted a series of RT experiments at UCLA from 2020 to 2022, including large-scale remote data-taking due to the COVID-19 pandemic. About one hundred undergraduate students in a breadth of STEM majors have participated in this research project, from protocol design to final data analysis. Over this period, we successfully recruited more than 200 auxiliary students as external unbiased participants. The detailed procedures results are available in our recent publications (Afifa et al. 2022; Bustanoby et al. 2022; Le et al. 2022; Ta et al. 2022)

Figures 27 - **29** summarize the remarkable results. First, **Figure 27** presents examples of the raw data distribution and analysis procedure presented in Afifa et al. 2022). The study consolidates SRT and CRT data from N = 40 participants in a protocol where one or two faces were displayed on a large (55") TV at randomly generated horizontal eccentricity (i.e., Azimuth) at 0, ±10^o, ±20^o, ±30^o, and ±40^o, at indiscriminate timings. **Figure 27-A** and **Figure 27-B** compile the raw data on covert attention and overt attention, respectively, with the solid black lines representing the aggregated data after global linear fitting by chi-square minimization. **Figure 27-C** includes the renormalized distribution of (A) after subtracting each participant's time offset, in order to evaluate the consistency of the slope across the data. **Figure 27-D** is the renormalized version of (B). Compelling linear relationships between SRT/CRT and the eccentricity are evident, exactly as predicted in **Figure 26**.

The following **Figures 28** and **29** summarize all aggregated plots of SRT or CRT as a function of various sets of 7D frame translations. **Figure 28** provides a summary of the CRT experiment for human facial recognition through scanning 5D phase space. **Figure 28-A** gives aggregated plots for the 1D space of Azimuth, as shown in **Figure 27** as well, including the recognition RT of one, two, and three faces. **Figure 28-B** displays the results of performing 3D face rotations by [Yaw, Pitch, Roll] (Le et al. 2022). Among the three rotational axes, transformations in the Roll axis translate into the flattest V-shape in data, as a result of the Log-polar primary visual cortex only having to process a simple 2D rotation (see **Section 3.2**). In contrast, processing transformations in Yaw and Pitch for the recognition of the true 3D location of the center of the 3D face is a far more complex endeavor (see **Section 6.5**). Nonetheless, variations in Yaw is shown to be easier to process than those in Pitch, perhaps as a result of Yaw transformations being more natural in our day-to-day applications of vision and navigation as compared to Pitch.

With respect to scaling, **Figure 28-C** illustrates clear scaling by Log(height of the face), as predicted in the discussion on the Log-polar coordinate configuration of the primary visual cortex (see **Section 3.2**). Even for faces memorized at 4^o, the fastest RT is recorded when the face is displayed again at 8^o, which indicates that we tend to normalize human faces as 8^o high objects as we memorize them (at to 1.4-m distance). Another remarkable discovery illustrated in **Figure 28-D** is that conscious perception of human faces seems to induce proper depth perception; such a task is realized through comparing the perceived size of the displayed face, against the memorized generically sized human face (with a height of ~20 cm). This adheres to our model in **Section 4.7**, even though the faces that were shown on the TV at a fixed distance were displayed at various sizes.

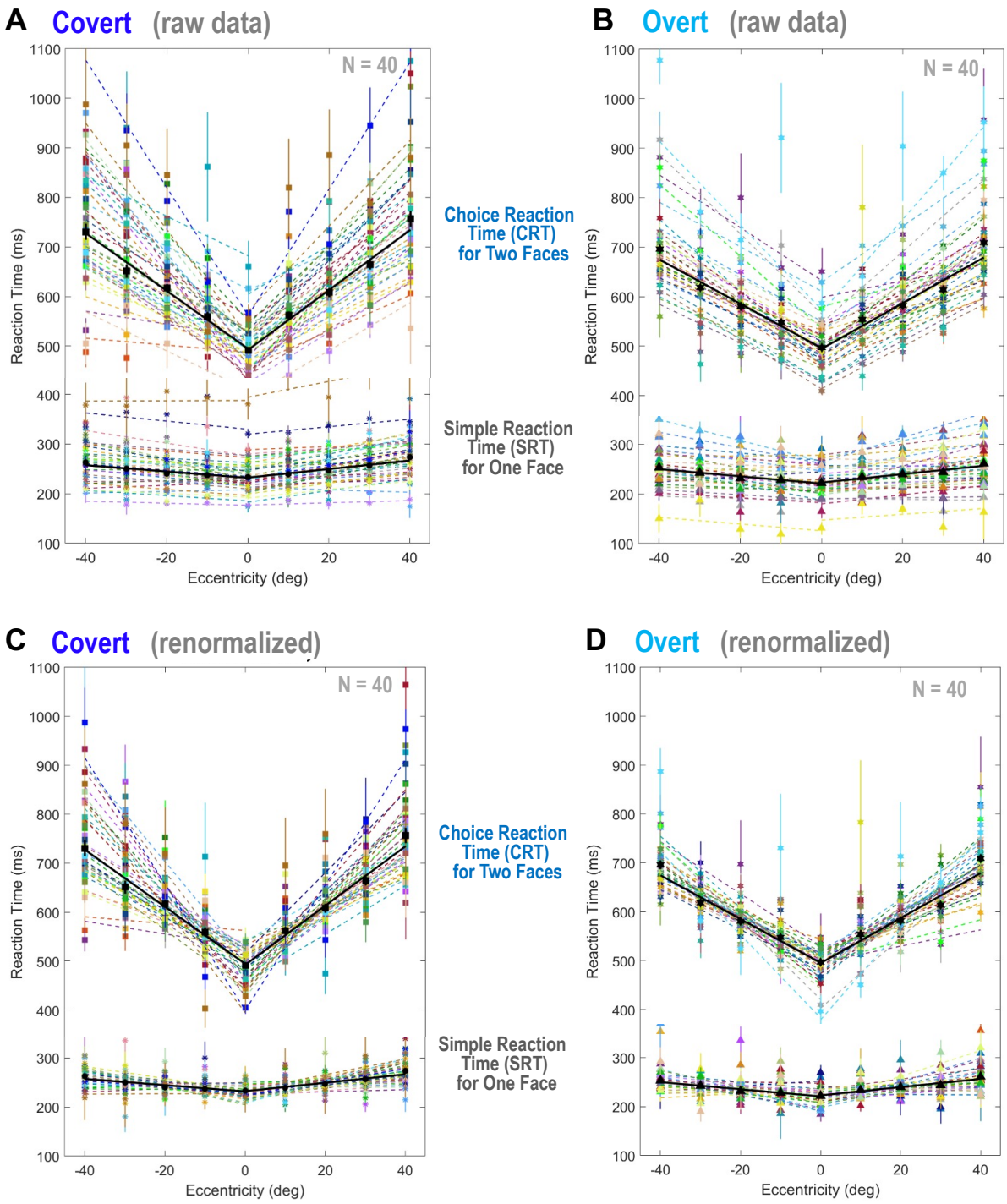

Figure 27. Simple and choice reaction times (SRT and CRT) of the recongition of one and two faces as a function of the horizontal eccentricity (i.e., Azimuth) for N = 40 participants. (The solid black lines indicate aggegated data.) **(A)** shows the raw data distribution of N = 40 in the covert attention study. **(B)** shows the raw data in the overt attention study. **(C)** specifies the renormalized distribution of (A) after time offset has been subtracted; this distribution helps us evaluate the consistency of the slope across the data collected. **(D)** is the renormalized version of (B). Compelling linear relationships between SRT/CRT and the eccentricity are evident, as predicted in **Figure 26**.

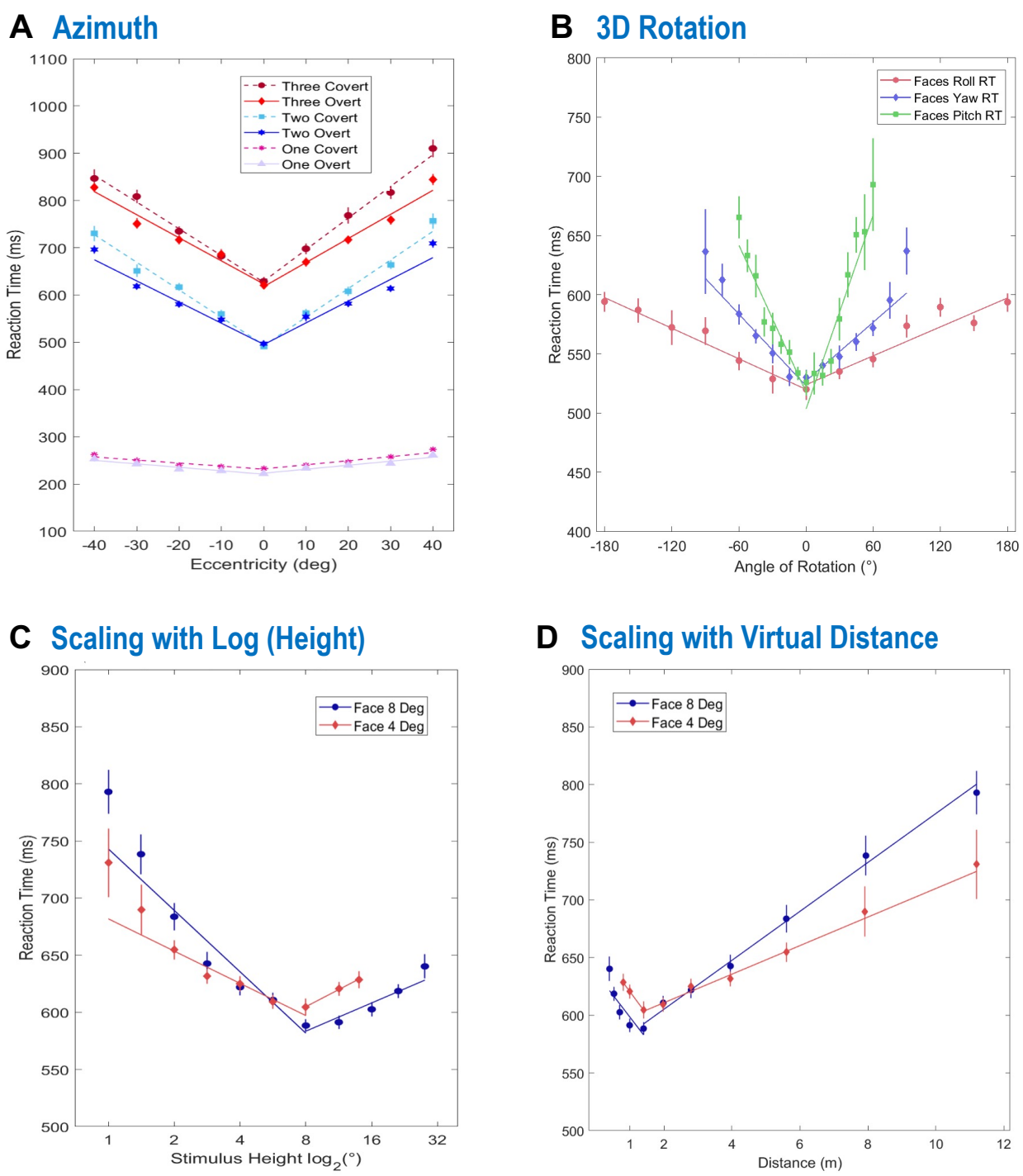

Figure 28. Observed Reaction Time vs. Scaling and 3D Rotation, taken as unfamiliar faces were shown on a large (50 -55 inch) TV. **(A)** plots CRT against Azimuth (i.e., Yaw) under the six conditions, including the display of 1, 2, 3, unfamiliar faces under covert and overt attention, respectively (Afifa et al. 2022). **(B)** plots CRT against rotations along Yaw, Pitch, and Roll axes(Le et al. 2022). **(C)** and **(D)** showcase CRT as a function of the faces' height (Ta et al. 2022). The horizontal axis of **(C)** represents Log (Height), whereas that of **(D)** indicates virtual linear distance (in meters) derived as 1.4 m x [8º / (Face height)].

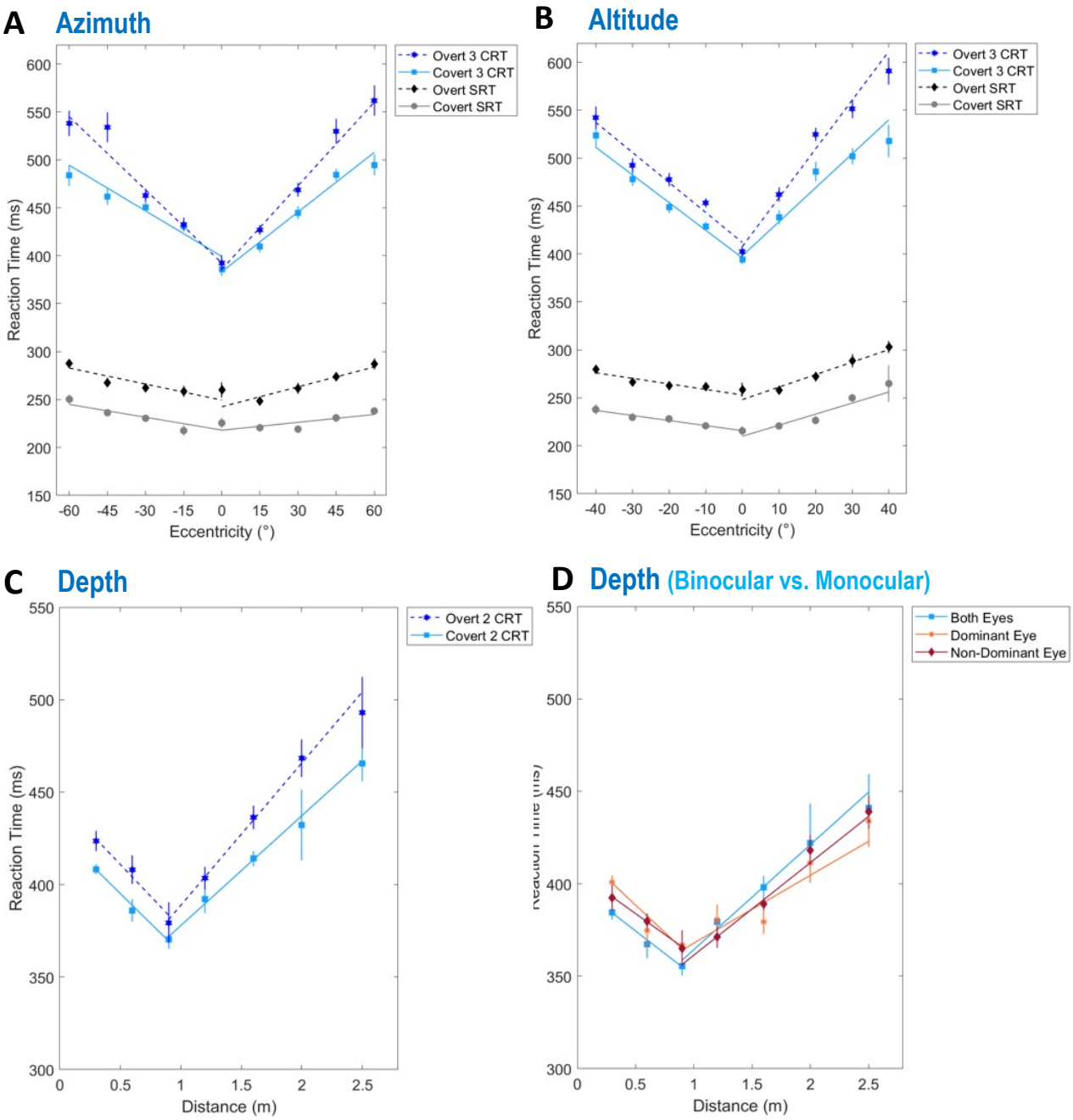

Figure 29. Observed Reaction Time vs. [Azimuth, Altitude, Depth] drawn from experiments where LED strips are placed along the three directions**. (A)** and **(B)** plots RT (in ms) against Azimuth (i.e., Yaw) and Altitude (i.e., Pitch) (both in degree), respectively (Afifa et al. 2022). Data were taken under four conditions: SRT involving one flashing LED strip, and CRT involving 1, 2, and 3 flashing LED strips, respectively; subjects were instructed so that these conditions were studied under both covert and overt attention. Saccadic eye movements were monitored by a high-speed IR CMOS camera to further verify subjects' covert and overt attention throughout the experiments. **(C)** and **(D)** plot RT (in ms) against linear distance (in meters) (Bustanoby et al. 2022). Data collected here include CRT when one or two flashing LED strips were displayed. **(C)** elucidates the difference between overt and covert attention. **(D)** shows data under three conditions: binocular, monocular with a dominant eye, monocular with a non-dominant eye, respectively.

Although depth perception cannot be directly studied with a large flat TV display, in principle, all of the other 6D can be systematically investigated; experiment design could utilize the distinctive presentation of various shapes, including human faces, Gabor patterns, colors, abstract patterns (e.g., triangle, square, etc.), and written language (familiar and unfamiliar characters, alphabets, and words). We have performed all of these experiments and obtained consistent supportive results with neat V-shaped patterns in all figures, as reported in the series of our recent publications (Afifa et al. 2022; Bustanoby et al. 2022; Le et al. 2022; Ta et al. 2022).

Studies on true 3D space and depth perception can be conducted with flashing LED strips placed in 3D space. **Figure 29** shows the results from such experiments that we have performed, in which flashing lights were generated by LED strips positioned along the three axes of [Azimuth, Altitude, Distance] (Afifa et al. 2022; Bustanoby et al. 2022). Our results affirmed that these LED strips are especially optimal for the study of depth perception. CRT data all showcased prominent V-shaped patterns, while flat horizontal lines were observed in SRT data, just as predicted in **Figure 26**.

In the case of the experiments involving LED strips, as shown in **Figure 29**, Overt RT appears to be slightly longer than Covert RT (i.e., perception is slower under overt attention), especially at the far periphery. This is opposite to the patterns observed in similar experiments conducted with the large TV, as displayed in **Figure 28-A**. These minor effects could explain for the slightly shaky vision that occurs after a substantial saccadic movement; however, this matter prompts future investigation.

Nevertheless, the prominent V-shaped patterns in CRT data for all of the 3 dimensions support our models and predictions in **Figure 26**. Depth perception (in meters) vs. RT (in ms) data, in particular, is perfectly linear, with the fastest RT being observed at initial cue location (at Z = 0.9 m). This initial cue location in 3D space becomes the center of the three axes in [Azimuth, Altitude, Distance] (i.e., [Yaw, Pitch, Distance], or [X, Y, Z]). Another notable observation here is that depth can be reliably perceived by either binocular or monocular vision, which supports our systematic analysis of depth perception in **Section 4**; most depth sensation comes from monocular views.

In conclusion, results from our systematic studies of reaction time strongly support the holographic theory of 3D vision by **MePMoS**, **NHT**, and **HAL** that we proposed. It would be impossible to explain these prominent V-shaped patterns in 7D phase spaces, as presented in **Figures 28** and **29**, without the basic principle of space-to-time conversion in 3D visual perception. With this theory in mind, one can conclude that fundamentally, we perceive 3D space by assigning information – converted from space to time – to each landmark within the external 3D allocentric space. Specifically, we generate the 3D visual perception internally (mentally), by projecting (predicting) the time information that is expected in three directions, similar to the process through which laser beams accomplish 3D projection mapping.

7.3 The Origin of Visual Illusions Explained

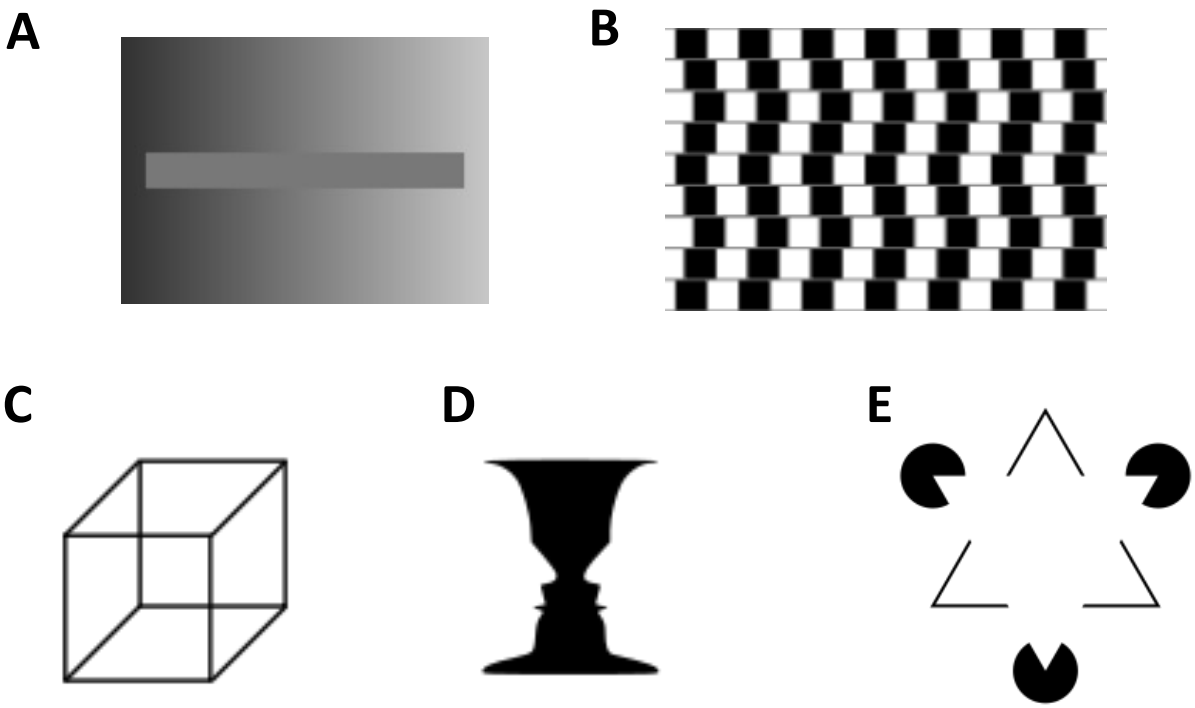

Figure 30. Examples of well-known optical illusions, for which our new model for the visual perception of 3D space and shapes are able to offer thorough explanations. **(A)** Although the bar in the middle is uniformly filled by the same grey colour, the right side of the bar appears to be darker. **(B)** The set of parallel horizontal lines appears to be tilted. **(C)** The Necker cube. **(D)** The Rubin vase, which can be perceived as an image of either two faces or a vase, simultaneously. **(E)** The Kanizsa triangle.

Our holographic model is able to provide ideal explanations for almost all known visual illusions, such as the most well-known ones included in **Figure 29**. The horizontal bar in **Figure 29-A** is uniformly filled with the same shade of grey, but it is perceived to be a gradient that spans in the opposite direction from that of the background gradient. Such an illusion occurs due to our inability to observe both the left side and the right side of the bar at the same time; only when the visual information (observed under either covert or overt attention) has been converted into time difference can the visual perception of the horizontal bar occur. Thus, we must presume the brightness along the horizontal bar as a function of time, with reference to brightness of the background; this results in the illusion of the horizontal bar being a gradient. It is for the same reason that **Figure 29-B** generates the illusion of the horizontal lines being alternatingly converging and diverging. The space-to time conversion process that is essential for perception results in our inability to perceive all the parallel lines at once. While our attention is shifted across the image, we extrapolate the impression of the these horizontal line segments being locally converging or diverging.

Optical illusions observed in **Figures 29-C**, **29-D**, and **29-E** can be explained through **MePMoS** and **NHT** as well. The illusions created by these figures all result from top-down signal processing in visual perception.

The Necker cube in **Figure 29-C** gives rise to the 3D sensations the same cube simultaneously facing different directions. This is due to our remarkable ability to generate depth perception from a 2D image via top-down higher-order cognitive processes (as discussed in **Section 4.8**). This 2D image is perfectly consistent with the representation of a 3D cube in multiple orientations, which can be generated in our 3D visual perception through prediction. Since our visual perception is an internal projection constructed via traveling alpha brainwaves, so long as the prediction agrees with the incoming 2D retinotopic image, the perception of 3D cube occurs and is in turn rewarded.

The Rubin vase illusion in **Figure 29-D** obeys the same principle of top-down processing through **MePMoS**. We can perceive either two white faces or a black vase, depending on our mind's prediction. The perceived image is then projected back to the actual 2D image.

The Kanizsa triangle in **Figure 29-E** appears consistent with the perception of a white inverted triangle in front of an upright triangle with black outline. Although not explicitly outlined, the white inverted triangle appears as an illusion enabled by the same top-down processing as in previous figures.

Similar to the figures discussed above, almost all existing displays used in magic take advantage of the simple space-to-time conversion process in visual perception. Since we are unable to pa attention to information at two separate locations simultaneously, as long as the magician manages to keep the audience's attention at one location, tricks can be executed at a different location.

Prism adaptation is another good example of where our theory applies (Luauté et al., 2009; Redding, Rossetti, and Wallace, 2005; Redding and Wallace, 2006). If we wear a prism that horizontally shifts the image, a participant can compensate for the shift to recover normal vision of external space rather quickly. Since our visual perception of the external world is purely the creation of internal prediction in time, if the PN shifts the phase of the alpha brainwave by the proper amount along the horizontal axis, then the perceived image is shifted back to normal. This can be effectively achieved through sensory-motor integration by ensuring that our eyes/head/body's appropriate alignment with visual inputs.

The required horizontal shift is easy to fix, but what about an inverted prism, when one wears upside-down goggles? It is known that people take a few weeks to get used to inverted images (Sachse et al., 2017). As outlined in our model, the inverted image can be flipped back by an appropriate switch of the alpha brainwave direction. If the image is inverted, the corollary discharge goes to the opposite vertical direction. Fixing this image requires the PN to swap the phase assignment of the alpha brainwaves to the opposite direction for self-consistency. Then, sensory-motor integration becomes normal, and visual perception in turn becomes normal as well.

It is also notable that our holographic theory is consistent with the old class of Gestalt psychology, which emphasizes that we perceive entire patterns, rather than separate individual parts. This is a result of our visual perception of external 3D space being a prediction-based, fundamentally top-down process.

8 Remaining Topics on 3D Vision

8.1 Local Shape and Color by Bottom-up Gamma Brainwaves

So far, we have emphasized that our visual perception is based on top-down processing based on prediction, which is the essence of **MePMoS** (**Me**mory → **P**rediction → **Mo**tion → **S**ensing). But this view is too simplistic and one-sided unless we incorporate it with bottom-up processes. Let us revisit the space-time diagram of **Figure 1** and **Figure 6-B** more closely. Bottom-up sensory signals from the eyes are processed by gamma brainwaves (40 – 100 Hz).

This bottom-up process by the gamma band has been well observed (Gray and Singer 1989; Fries et al. 2001, 2007, 2008; Vinck et al. 2010; Brunet et al. 2013). We already utilized the gamma brainwaves for local depth sensation on V1-V3 in **Section 4.3**. Besides local depth perception, the gamma bands should take care of various kinds of local visual information processing including the direction of local line segments, contrasts, color, and brightness.

Indeed, such local signal processing begins just after the photoreceptors on the retina by the horizontal cells (for contrast enhancement) and by the bipolar cells (for time differentiation). The primary visual cortex, V1-V3, continues to process local shape and color from the bottom up system in parallel. This is the traditional concept of enlarging Receptive Fields (RF) through this bottom-up processing.

The handshake of the bottom-up and top-down processes is essential for the recognition of complex objects such as human faces. We ought to process the shape and color of eyes, lips, nose, and hair in an extremely efficient way in parallel. We will discuss this critical process in **Part V** when we introduce the concept of the **Grand Unification** of the five senses. Roughly speaking, all five senses are expressed and encoded, respectively, by high-frequency gamma bands (40 – 100 Hz). There are localized, specialized processes running in parallel. Our perception of global semantic 3D shape and 3D space is the outcome of the handshake between the top-down and bottom-up streams given in **Figure 1-A**.

8.2 Origin of Superior Visual Accuity

Another concern or shortcoming of the proposed **3D Vision HAL** is the lack of superior visual acuity, which we addressed once in **Part II: Section 7.2** when we introduced the generic concept of **HAL**. Since the effect of poor resolution seems the most severe in vision, let us examine it again.

Our vision possesses excellent visual acuity of 20/20 achieved by ~100 million photoreceptors. It seems even better than the 4K resolution of ~2,000 x 2,000 = four million pixels. On the other hand, the proposed **HAL** has a lattice structure with only 16 neurons per string assigned to each dimension, which is way too short to explain our demonstrated visual acuity.

The fundamental limitation of spatial resolution by the **HAL** is due to space-to-time conversion utilizing the phase information of alpha brainwaves. Generally speaking, 1/10 of the temporal resolution of the period is feasible by assigning a specific phase, which limits the number of neurons per string to an order of ten; thus, 16 neurons is already on the optimistic side. It also coincides with the observed time resolution of gamma phase measurement. Gray and Singer reported the gamma phase correlation to spike timing in the orientation column of the cat visual cortex with (0.4 ± 1.9)-millisecond accuracy in 1989 (Gray & Singer, 1989). A similar temporal resolution was reported by studying Monkey visual cortex (Fries, Nikolić, & Singer, 2007; Vinck, Womelsdorf, & Fries, 2013). These papers support the 16 segments of **HAL** per one period of various brainwaves below.

- Theta wave (5 Hz) → 200 ms / 16 ~ 12 ms
- Alpha wave (10 Hz) → 100 ms / 16 ~ 7 ms
- Beta wave (20 Hz) → 50 ms / 16 ~ 3 ms
- Gamma wave (50 Hz) → 20 ms / 16 ~ 1 ms

How then can we improve the effective temporal resolution in time from 1/16 up to the order of 1/1,000? To remedy this challenge, we can consider the following four major contributions to improving visual acuity, where (1) is specific to human vision.

1) The Log-polar coordinate system of V1.
2) Micro-saccades
3) Detailed local shape processed by bottom-up gamma brainwaves.
4) Discrete Fourie Transformation (DFT) by alpha brainwaves.

Firstly, the log-polar coordinate system dramatically improves visual acuity toward the foveal center, as illustrated by the mapping of a human face from the retina to V1 in **Figure 9**. Our visual acuity of 20/20 is only true at the exact dead center of the fovea. The actual acuity deteriorates dramatically towards the peripheral region. As shown in **Section 5.3**, covert attention only allows us to read ~10 alphabetic characters at once (within the red circle.). So-called 4K resolution is a kind of visual illusion, supported by hidden unconscious overt attention, which reconstructs allocentric (body-centric) high-resolution images, piece by piece, as a function of time.

Secondly, details of the image at the foveal center can be further improved by rhythmic, predicable micro-saccades. By taking advantage of **MePMoS**, fine spatial details can be transformed to the time domain by micro-saccades, as shown in **Section 4.5: Figure 15**. Since our visual perception is fundamentally in the time domain, the micro-saccades – mapped on the Log-polar V1 and scanned by the constant-speed alpha waves – could contribute the most in superior visual acuity at the foveal center. By integrating with binocular disparity covered in **Section 4.4**, human vision can reproduce extremely detailed 3D shape accurately, as we experience.

Thirdly, local patterns and shapes are processed by bottom-up gamma waves in parallel, which creates the semantic sensation of the local shapes, such as eyes and lips for example, in far more detail than top-down alpha-only vision. Let's consider the linkage of **Vision HAL** by alpha (10 Hz) linked with a **Gamma Shape HAL** (~100 Hz). Gamma waves, which are higher in frequency by a factor of ten, effectively work as a 10x magnifier for local regions of interest (ROI). The bottom-up gamma process particularly helps to read and understand written language at high speed, which will be discussed in **Part VI**.

Lastly, in the frequency-time domain, there is a well-known mathematical treatment to enhance resolution arbitrarily well, called Fourie transformation which fundamentally follows space-to-time conversion. Brains seem to perform Discrete Fourie Transformation (DFT), which appears to be observed by the discrete pattern of grid cells achieved by theta brainwaves in the Hippocampal network for navigation (Stensola et al. 2012). We will discuss this at length in the following **Part IV**. If that is the case for improving the accuracy of navigation by theta waves, our vision must have utilized the similar DFT by alpha brainwaves.

In summary, the combination of the above four contributions should be able to achieve our superior visual acuity at the foveal center. Hopefully, more quantitative analysis and numerical models will clarify this question in the near future.

8.3 Evolution of Vision

Animal Kind	Navigation or Vision	Dimen-sion	Cartesian vs. Polar	Linear vs. Log	Axes	Remarks	Proved ?
Hydra, Jelly fish							
	Navigation	1D	Polar	Linear	[Yaw]	Difused Nerve Ring	Yes
C. elegans							
	Navigation	2D	Cartesian	Linear	[X, Y]	CPG based	Yes
Insect							
	Vision	2D	Polar	Linear	[Yaw, Pitch]	2D image on Retina	Yes
		3D	Polar	Linear	[Yaw, Pitch, Distance]	2D Ring Attractor	No
	Navigation	2D	Cartesian	Linear	[X', Y']	1D Ring, 45° tilted	Yes
		3D	Cartesian	Linear	[X', Y', Z]	2D Ring, 45° tilted	No
Bird							
	Vision	2D	Polar	Linear	[Yaw, Pitch]	2D image on Retina	Yes
		2D	Polar	Linear	[Yaw, Pitch]	Primary Visual Cortex	Yes
	Navigation	3D	Cartesian	Linear	[X', Y', Z] 45° tilted	MEC - Hippocampus	No
Rodent							
	Vision	2D	Polar	Linear	[Yaw, Pitch]	2D image on Retina	Yes
		2D	Polar	Linear	[Yaw, Pitch]	Primary Visual Cortex	Yes
	Navigation	2D	Cartesian	Linear	[X", Y"] Hex grid	MEC - Hippocampus	Yes
		3D	Cartesian	Linear	[X", Y", Z"] Hex tilted	MEC - Hippocampus	No
Human							
	Vision	2D	Polar	Linear	[Yaw, Pitch]	2D image on Retina	Yes
		2D	Polar	Log	[Roll, Log(Eccentricity)]	Primary Visual Cortex	Yes
		2D	Polar	Linear	[Roll, Eccentricity]	Dorsal (MT-FEF)	No
			Polar	Linear	[Yaw, Pitch, Roll, Distance]	Dorsal (FEF-7a)	Yes
		3D	Polar	Log	[Yaw, Pitch, Roll, Log(Dis.)]	Ventral (VTC)	No
	Navigation	2D	Cartesian	Linear	[X", Y"] Hex grid	Entorhinal Cortex	Yes
		3D	Cartesian	Linear	[X", Y", Z"] Hex tilted	Entorhinal Cortex	No
		3D	Polar	Linear	[Yaw, Pitch, Roll, Distance]	Parahippocampal Cortex	No
		3D	Polar	Log	[Yaw, Pitch, Roll, Log(Dis.)]	Retrosplenial Cortex	No

Table 3. Summary table of coordinate systems of various animals. From top to bottom, it follows evolutionary developmental stages. (This is taken from **Part II: Section 4.2, Table 2**.)

Through this **Part III**, we have discussed and compared the various visual systems of different animals – from insects, birds, rodents, and monkeys – to clarify the unique feature of human vision. The evolution of visual acuity is systematically reviewed by Caves et al (2018). Here, we shall revisit and focus on the evolution of the coordinate systems (covered in **Section 3.3**) and depth perception (covered **in Section 4.10**.) from a global perspective. To begin with, all the coordinate systems for both vision and navigation are listed in **Table 3** (taken from **Part II: Section 4.2, Table 2**.)

From the top row to the bottom, **Table 3** follows the evolutionary steps in visual processes in order of complexity as found in extant taxa going from Hydra/Jellyfish, to *C. elegans*, to Insects and other arthropods, to birds and other non-mammalian vertebrates, to rodents and other non-primate mammals, to monkeys, and finally to humans. From an evolutionary point of view, navigation systems emerged first in stem chordates, and were likely similar to extant *C. elegans* (~800 million years ago), prior to sophisticated eyes and accompanying visual systems as found in extant arthropods such as insects (~550 million years ago). In between, we could consider primitive visual systems like a pair of light sensors of snails, or 16 light sensors of scallops (Phylum Mollusca). By the time insects and other arthropods emerged on the Earth, true eyes for 2D vision with a few thousand individual sensors as compound eyes had evolved.

From the table, we can derive a few clear conclusions. Firstly, navigation has always been based on the linear Cartesian coordinate systems of [X, Y, Z] from the beginning similar to some of the most primitive extant phyla (e.g., *C. elegans*) to the most evolutionarily sophisticated visual system found in humans, simply because navigation should be fundamentally linear translations in either 2D or 3D space. In contrast, vision evolved and has been conserved under the polar coordinate system of [Yaw, Pitch, Distance], because light must be bombarded on the eyes through straight lines; Thus, projection back to the emitted points must be in polar coordinates.

In both cases, we believe that both were 3D rather than 2D from the evolutionary origin. It is because, in our **NHT** and **HAL**, both 2D space and 3D space are compressed to a similar lattice structure anyway. So, the extension from 2D to 3D is trivial. Especially, insects and birds fly in the 3D atmosphere, and fishes swim in the 3D aqueous medium. Therefore, it is natural to assume that all other animals inherited their 3D navigation and vision systems from these flying or swimming animals.

So far, all animals share identical coordinates: the Cartesian for navigation and the Polar for vision. But then, there is one striking fact that differentiates humans from all others. That is, our primary visual cortex is the Log-polar coordinate of [Roll, Log(Eccentricity)]. Even monkeys have not quite developed the Log-polar V1. It seems halfway from the Cartesian to Log-polar (Vanduffel et al. 2014). Birds and rodents have the perfect Linear-polar visual cortex of [Yaw, Pitch].

Through our extensive investigation on this issue, we have identified three major merits of the Log-polar visual cortex.

1) Prompt and accurate scaling and Roll rotation to recognize semantic shape.
2) Depth perception to longer distance (>> 10 m).
3) Reliable pursuit of moving targets (like prey.)

In conclusion, human vision seems specialized for visual perception of 3D space and semantic shape.

8.4 What is Vision? Remaining Questions

Through this **Part III**, we have expanded the new concept that space must be converted to time by brainwaves, following **NHT** and **HAL** in **Part II**. The critical conclusion is that we cannot perceptive space unless we predict and project it like a 3D projection mapping. Conscious awareness of visual perception is fundamentally a top-down creation of our mind. To some extent, it is like a dream that must be rewarded by the reality of visual stimulation. In this scheme, there is no binding problem at all.

Since it is an internal creation, we can easily imagine and generate the visual scene by 3D – as we wish – purely based on 2D retinotopic images. Furthermore, our perception of space could also have dual coordinate systems: One for strict vision and the other for navigation purposes. The former is

based on the body-centric Linear-polar coordinate system organized by alpha brainwaves (~10 Hz), and the latter is by the allocentric Cartesian coordinate system by theta brainwaves (~5 Hz).

We have paid special attention to the peculiar Log-polar coordinate system of our primary visual cortex, which is so unique to the human. The Log-polar system dramatically helps to enhance our predictive power of the future for several reasons:

1) We can dramatically improve visual acuity at the foveal center.
2) We can recognize 2D shapes by scaling and rotation promptly and reliably.
3) We can sense and estimate distance far away (>> 10 m)

The primary purpose of the brain is a prediction for better navigation in the future. From this point of view, the Log-polar visual system appears the ultimate machine by nature to achieve this goal.

Of course, a new answer creates more questions in science. Most notably, we consider the following directions particularly important and promising.

1) Direct detection of EEG signals along the visual cortex to observe the direct correlation between the RF by visual stimulation and the phase of the alpha waves.
2) Single-neuron-level research of insect visual systems: Is **HNT** and **HAL** applicable?
3) Direct evidence of the conversion of the coordinate system.

In the next **Part IV**, we shall move on to the navigation system by the Hippocampal network. Then, we can finally define the origin of memories, the starting point of **MePMoS**.

Contributions

KA developed the new concept and theory of the holographic 3D vision by applying **MeMoS**, **NHT,** and **HAL** from **Part I** and **Part II**. He drafted the manuscript with all the figures and tables. AB contributed to fine-tuning the holographic concept, in particular, from his expertise on the visual systems of birds and many other animals. He also revised and polished the manuscript.

Acknowledgments

KA thanks Syed Hydari for his contribution to the early stage of concept development. Discussion with Zahra Aghajan was also helpful to confirm new ideas from time to time. We thank Ziyan Peng and Roshan Gunturu for their careful editing of the manuscript, and Justin Yi for the useful conversation. Special thanks go to the three graduate students for their hard work to design and operate our new labs: Javier Carmona, Chandan Kittur, and Elizabeth Mills.

In addition, numerous undergraduate students at UCLA participated in the experimental projects described in **Section 7.1**. All the experiments have been conducted at UCLA as a part of the Elegant Mind Club (EMC), either locally or remotely during the pandemic from 2020 to 2022. During this period, more than a hundred students participated in every phase of the experiments.

To name a few, the following made indispensable contributions: Umaima Afifa, Patrick Wilson, Brian Ta, Amy Dinh, Diego Espino, Chris Dao, Saba Doust, Trevor McCarthy, Natsuko Yamaguchi, Isabella Bustanoby, Andrew Krupien, Maria Eduarda Mendes Silva, Caominh Le, Nathaniel Chen, Christina Honore, Anthony Garibay, Despoina Sparakis, April Smith, Alyssa Drost, Sukanya Mohapatra, Jonathan Chan, Megan Yu, Jagannathan Rangarajan, Kimya Peyvan, Mira Khosla, KyleTsujimoto, Benjamin Asdell, Mark Diamond, Kelly Bartlett, Vedang Bhelande, Kailey Fleiszig-Evens, Tim Van Hoomissen, Erica Li, Gurleen Kaur, Felicia Wang, Kelly Bartlett, Meera McAdam, Jared Khoo, Cindy Ta, Leonard Schummer, Angela East, Amanda Yares, Jinwoo Baik, Kenya Ochoa, Michaela Bacani, Alice Yanovsky, Erica Li, Shevin Nia, and many others.

Furthermore, a total of ~500 extra students helped us as participants in various stages of RT experiments in 2020 – 2022. Especially, about 200 students volunteered as unbiased participants for the systematic final data taking in Fall 2021. We thank them for their kind commitment.

This work was in part supported by the Dean's office of life science, Dean's office of physical science, Chair's office of the department of physics and astronomy, and the Instructional Improvement grant by the Center for the Advancement of Teaching, all at the University of California, Los Angeles.

References

Abdollahi, Rouhollah O., Hauke Kolster, Matthew F. Glasser, Emma C. Robinson, Timothy S. Coalson, Donna Dierker, Mark Jenkinson, David C. Van Essen, and Guy A. Orban. 2014. "Correspondences between Retinotopic Areas and Myelin Maps in Human Visual Cortex." *NeuroImage* 99:509–24. doi: 10.1016/j.neuroimage.2014.06.042.

Adesnik, Hillel, William Bruns, Hiroki Taniguchi, Z. Josh Huang, and Massimo Scanziani. 2012. "A Neural Circuit for Spatial Summation in Visual Cortex." *Nature* 490(7419):226–31. doi: 10.1038/nature11526.

Afifa, Umaima, Javier Carmona, Amy Dinh, Diego Espino, Trevor McCarthy, Brian Ta, Patrick Wilson, Benjamin Asdell, Jinwoo Baik, Archana Biju, Sonia Chung, Christopher Dao, Mark Diamond, Saba Doustmohammadi, Angela East, Kailey Fleiszig-Evans, Adrian Franco, Anthony Garibay-Gutierrez, Aparajeeta Guha, Roshan Gunturu, Luke Handley, Christina Honore, Abinav Kannan, Jared Khoo, Mira Khosla, Chandan Kittur, Alexandra Kwon, Jessica Lee, Nicholas Lwe, Mylan Mayer, Elizabeth Mills, Delilah Pineda, Pasha Pourebrahim, Jacob Rajacich, Shan Rizvi, Liliana Rosales, Leonard Schummer, Conor Sefkow, Alexander Stangel, Cindy Ta, Ivy Ta, Natalie Tong, Kyle Tsujimoto, Alyssa Vu, Henry Wang, Amanda Yares, Natsuko Yamaguchi, Ki Woong Yoon, Shuyi Yu, Aaron P. Blaisdell, and Katsushi Arisaka. 2022. “Visual Perception of 3D Space and Shape in Time - Part I: 2D Space Perception by 2D Linear Translation.” 2022.03.01.482161.

Apitz, Holger, and Iris Salecker. 2014. “A Challenge of Numbers and Diversity: Neurogenesis in the *Drosophila* Optic Lobe.” *Journal of Neurogenetics* 28(3–4):233–49. doi: 10.3109/01677063.2014.922558.

Araujo, H., and J. M. Dias. 1996. “An Introduction to the Log-Polar Mapping [Image Sampling].” Pp. 139–44 in *Proceedings II Workshop on Cybernetic Vision*.

Arisaka, Katsushi. 2022a. “Grand Unified Theory of Mind and Brain - Part I: Space-Time Approach to Dynamic Connectomes of C. Elegans and Human Brains by MePMoS.”

Arisaka, Katsushi. 2022b. “Grand Unified Theory of Mind and Brain - Part II: Neural Holographic Tomography (NHT) and Holographic Ring Attractor Lattice (HAL).”

Benson, Noah C., Omar H. Butt, David H. Brainard, and Geoffrey K. Aguirre. 2014. “Correction of Distortion in Flattened Representations of the Cortical Surface Allows Prediction of V1-V3 Functional Organization from Anatomy” edited by W. Einhäuser. *PLoS Computational Biology* 10(3):e1003538. doi: 10.1371/journal.pcbi.1003538.

Bischof, Hans-Joachim, Dennis Eckmeier, Nina Keary, Siegrid Löwel, Uwe Mayer, and Neethu Michael. 2016. “Multiple Visual Field Representations in the Visual Wulst of a Laterally Eyed Bird, the Zebra Finch (Taeniopygia Guttata)” edited by J. J. Bolhuis. *PLOS ONE* 11(5):e0154927. doi: 10.1371/journal.pone.0154927.

Bridge, Holly, David A. Leopold, and James A. Bourne. 2016. “Adaptive Pulvinar Circuitry Supports Visual Cognition.” *Trends in Cognitive Sciences* 20(2):146–57. doi: 10.1016/j.tics.2015.10.003.

Brunet, N., C. A. Bosman, M. Roberts, R. Oostenveld, T. Womelsdorf, P. De Weerd, and P. Fries. 2013. “Visual Cortical Gamma-Band Activity During Free Viewing of Natural Images.” *Cerebral Cortex* 25(4):918–26.

Bustanoby, Isabella, Andrew Krupien, Umaima Afifa, Benjamin Asdell, Michaela Bacani, James Boudreau, Javier Carmona, Pranav Chandrashekar, Mark Diamond, Diego Espino, Arnav Gangal, Chandan Kittur, Yaochi Li, Tanvir Mann, Christian Matamoros, Trevor McCarthy, Elizabeth Mills, Stephen Nazareth, Justin Nguyen, Kenya Ochoa, Sophie Robbins, Despoina Sparakis, Brian Ta, Kian Trengove, Tyler Xu, Natsuko Yamaguchi, Christine Yang, Eden Zafran, Aaron P. Blaisdell, and Katsushi Arisaka. 2022. “Visual Perception of 3D Space and Shape in Time - Part II: 3D Space Perception with Holographic Depth.” 2022.02.28.482181.

Caves, Eleanor M., Nicholas C. Brandley, and Sönke Johnsen. 2018. “Visual Acuity and the Evolution of Signals.” *Trends in Ecology & Evolution* 33(5):358–72. doi: 10.1016/j.tree.2018.03.001.

Chauhan, Tushar, Yseult Héjja-Brichard, and Benoit R. Cottereau. 2020. “Modelling Binocular Disparity Processing from Statistics in Natural Scenes.” *Vision Research* 176:27–39. doi: 10.1016/j.visres.2020.07.009.

Coulon, Philippe, and Carole E. Landisman. 2017. “The Potential Role of Gap Junctional Plasticity in the Regulation of State.” *Neuron* 93(6):1275–95. doi: 10.1016/j.neuron.2017.02.041.

Cumming, B. G., and G. C. DeAngelis. 2001. “The Physiology of Stereopsis.” *Annual Review of Neuroscience* 24(1):203–38. doi: 10.1146/annurev.neuro.24.1.203.

DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust. 2012. “How Does the Brain Solve Visual Object Recognition?” *Neuron* 73(3):415–34. doi: 10.1016/j.neuron.2012.01.010.

Engbert, Ralf. 2006. “Microsaccades: A Microcosm for Research on Oculomotor Control, Attention, and Visual Perception.” Pp. 177–92 in *Progress in Brain Research*. Vol. 154, *Visual Perception*, edited by S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J.-M. Alonso, and P. U. Tse. Elsevier.

Felleman, Daniel J., and David C. Van Essen. 1991. “Distributed Hierarchical Processing in the Primate Cerebral Cortex.” *Cereb Cortex* 1–47.

Freedman, David J., and Guilhem Ibos. 2018. “An Integrative Framework for Sensory, Motor, and Cognitive Functions of the Posterior Parietal Cortex.” *Neuron* 97(6):1219–34. doi: 10.1016/j.neuron.2018.01.044.

Fries, P., T. Womelsdorf, R. Oostenveld, and R. Desimone. 2008. “The Effects of Visual Stimulation and Selective Visual Attention on Rhythmic Neuronal Synchronization in Macaque Area V4.” *Journal of Neuroscience* 28(18):4823–35. doi: 10.1523/JNEUROSCI.4499-07.2008.

Fries, Pascal, Danko Nikolić, and Wolf Singer. 2007. “The Gamma Cycle.” *Trends in Neurosciences* 30(7):309–16. doi: 10.1016/j.tins.2007.05.005.

Fries, Pascal, John H. Reynolds, Alan E. Rorie, and Robert Desimone. 2001. “Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention.” *Science* 291(5508):1560–63. doi: 10.1126/science.1055465.

Garrett, M. E., I. Nauhaus, J. H. Marshel, and E. M. Callaway. 2014. “Topography and Areal Organization of Mouse Visual Cortex.” *Journal of Neuroscience* 34(37):12587–600. doi: 10.1523/JNEUROSCI.1124-14.2014.

Gehring, Walter J. 2014. “The Evolution of Vision.” *WIREs Developmental Biology* 3(1):1–40. doi: 10.1002/wdev.96.

Gilbert, Charles D., and Wu Li. 2013. “Top-down Influences on Visual Processing.” *Nature Reviews Neuroscience* 14(5):350–63. doi: 10.1038/nrn3476.

Gray, C. M., and W. Singer. 1989. “Stimulus-Specific Neuronal Oscillations in Orientation Columns of Cat Visual Cortex.” *Proceedings of the National Academy of Sciences* 86(5):1698–1702. doi: 10.1073/pnas.86.5.1698.

Horton, Jonathan C., and William F. Hoyt. 1991. “The Representation of the Visual Field in Human Striate Cortex: A Revision of the Classic Holmes Map.” *Archives of Ophthalmology* 109(6):816–24. doi: 10.1001/archopht.1991.01080060080030.

Hubel, D. H., and T. N. Wiesel. 1959. “Receptive Fields of Single Neurones in the Cat’s Striate Cortex.” *The Journal of Physiology* 148(3):574–91. doi: 10.1113/jphysiol.1959.sp006308.

Hubel, D. H., and T. N. Wiesel. 1962. “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex.” *The Journal of Physiology* 160(1):106–54. doi: 10.1113/jphysiol.1962.sp006837.

Hubel, D. H., and T. N. Wiesel. 1968. “Receptive Fields and Functional Architecture of Monkey Striate Cortex.” *The Journal of Physiology* 195(1):215–43. doi: 10.1113/jphysiol.1968.sp008455.

Hubel, David H., and Torsten N. Wiesel. 1998. “Early Exploration of the Visual Cortex.” *Neuron* 20(3):401–12. doi: 10.1016/S0896-6273(00)80984-8.

Javier Traver, V., and Alexandre Bernardino. 2010. “A Review of Log-Polar Imaging for Visual Perception in Robotics.” *Robotics and Autonomous Systems* 58(4):378–98. doi: 10.1016/j.robot.2009.10.002.

Kim, HyunGoo R., Dora E. Angelaki, and Gregory C. DeAngelis. 2016. “The Neural Basis of Depth Perception from Motion Parallax.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 371(1697):20150256. doi: 10.1098/rstb.2015.0256.

Kolodkin, Alex L., and P. Robin Hiesinger. 2017. “Wiring Visual Systems: Common and Divergent Mechanisms and Principles.” *Current Opinion in Neurobiology* 42:128–35. doi: 10.1016/j.conb.2016.12.006.

Kral, Karl. 2003. “Behavioural–Analytical Studies of the Role of Head Movements in Depth Perception in Insects, Birds and Mammals.” *Behavioural Processes* 64(1):1–12. doi: 10.1016/S0376-6357(03)00054-8.

Kruger, N., P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez, and L. Wiskott. 2013. “Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1847–71. doi: 10.1109/TPAMI.2012.272.

Lappin, Joseph S., and Warren D. Craft. 1997. “Definition and Detection of Binocular Disparity.” *Vision Research* 37(21):2953–74. doi: 10.1016/S0042-6989(97)00091-6.

Laubrock, Jochen, Ralf Engbert, and Reinhold Kliegl. 2005. “Microsaccade Dynamics during Covert Attention.” *Vision Research* 45(6):721–30. doi: 10.1016/j.visres.2004.09.029.

Le, Caominh T., Samantha Pedersen, Jonathan Chan, Nathaniel Chen, Brian Ta, Patrick Wilson, Trevor McCarthy, Emma Barseghyan, Anushka Chauhan, Hind Saif, Jonathan Tu, Darren J. Wijaya, Annika Zhang, Erica Li, Camille Marangi, Setayesh Nekarae, Felicia Wang, Alice Yanovsky, Umaima Afifa, Javier Carmona, Diego Espino, Leonard Schummer, Philip A. Gudijanto, Gurleen Kaur, Andrew Lam, Matthew Mar, Elizabeth Mills, Alexandra Nevins, Elijah Ortiz, Kyle Wheeler, Aaron Blaisdell, and Katsushi Arisaka. 2022. “Visual Perception of 3D Space and Shape In Time - Part IV: 3D Shape Recognition by 3D Rotation.” 2022.03.01.482164.

Llinas, Rodolfo R. 2014. “Intrinsic Electrical Properties of Mammalian Neurons and CNS Function: A Historical Perspective.” *Frontiers in Cellular Neuroscience* 8. doi: 10.3389/fncel.2014.00320.

Lozano-Soldevilla, Diego, and Rufin VanRullen. 2019. “The Hidden Spatial Dimension of Alpha: 10-Hz Perceptual Echoes Propagate as Periodic Traveling Waves in the Human Brain.” *Cell Reports* 26(2):374-380.e4. doi: 10.1016/j.celrep.2018.12.058.

Lyu, Cheng, L. F. Abbott, and Gaby Maimon. 2020. *A Neuronal Circuit for Vector Computation Builds an Allocentric Traveling-Direction Signal in the* Drosophila *Fan-Shaped Body*. *preprint*. Neuroscience. doi: 10.1101/2020.12.22.423967.

Martinez-Conde, Susana, Jorge Otero-Millan, and Stephen L. Macknik. 2013. “The Impact of Microsaccades on Vision: Towards a Unified Theory of Saccadic Function.” *Nature Reviews Neuroscience* 14:83–96.

Mauss, Alex S., Katarina Pankova, Alexander Arenz, Aljoscha Nern, Gerald M. Rubin, and Alexander Borst. 2015. “Neural Circuit to Integrate Opposing Motions in the Visual Field.” *Cell* 162(2):351–62. doi: 10.1016/j.cell.2015.06.035.

Melloni, Lucia, Caspar M. Schwiedrzik, Eugenio Rodriguez, and Wolf Singer. 2009. “(Micro)Saccades, Corollary Activity and Cortical Oscillations.” *Trends in Cognitive Sciences* 13(6):239–45. doi: 10.1016/j.tics.2009.03.007.

Miller, Earl K., and Timothy J. Buschman. 2013. “Cortical Circuits for the Control of Attention.” *Current Opinion in Neurobiology* 23(2):216–22. doi: 10.1016/j.conb.2012.11.011.

Morris, A. P., F. Bremmer, and B. Krekelberg. 2013. “Eye-Position Signals in the Dorsal Visual System Are Accurate and Precise on Short Timescales.” *Journal of Neuroscience* 33(30):12395–406. doi: 10.1523/JNEUROSCI.0576-13.2013.

Morris, Adam P., Michael Kubischik, Klaus-Peter Hoffmann, Bart Krekelberg, and Frank Bremmer. 2012. “Dynamics of Eye-Position Signals in the Dorsal Visual System.” *Current Biology* 22(3):173–79. doi: 10.1016/j.cub.2011.12.032.

Nadler, Jacob W., Dora E. Angelaki, and Gregory C. DeAngelis. 2008. “A Neural Representation of Depth from Motion Parallax in Macaque Visual Cortex.” *Nature* 452(7187):642–45. doi: 10.1038/nature06814.

Nériec, Nathalie, and Claude Desplan. 2016. “From the Eye to the Brain.” Pp. 247–71 in *Current Topics in Developmental Biology*. Vol. 116. Elsevier.

Nityananda, Vivek, and Jenny Read. 2017. “Stereopsis in Animals: Evolution, Function and Mechanisms | Journal of Experimental Biology | The Company of Biologists.” Retrieved January 8, 2022 (https://journals.biologists.com/jeb/article/220/14/2502/18621/Stereopsis-in-animals-evolution-function-and).

Okajima, Kenji. 2004. “Binocular Disparity Encoding Cells Generated through an Infomax Based Learning Algorithm.” *Neural Networks* 17(7):953–62. doi: 10.1016/j.neunet.2004.02.004.

Otsuna, Hideo, Kazunori Shinomiya, and Kei Ito. 2014. “Parallel Neural Pathways in Higher Visual Centers of the Drosophila Brain That Mediate Wavelength-Specific Behavior.” *Frontiers in Neural Circuits* 8. doi: 10.3389/fncir.2014.00008.

Patel, Gaurav H., Gordon L. Shulman, Justin T. Baker, Erbil Akbudak, Abraham Z. Snyder, Lawrence H. Snyder, and Maurizio Corbetta. 2010. “Topographic Organization of Macaque Area LIP.” *Proceedings of the National Academy of Sciences* 107(10):4728–33. doi: 10.1073/pnas.0908092107.

Pierrot-Deseilligny, Charles, Sophie Rivaud, Bertrand Gaymard, René Müri, and Anne-Isabelle Vermersch. 1995. “Cortical Control of Saccades.” *Annals of Neurology* 37(5):557–67. doi: https://doi.org/10.1002/ana.410370504.

Qi, Bin, and Takayuki Nakata. 2006. “A Log-Polar Transformation Method for Face Recognition.” 6.

Railo, Henry, Joni Saastamoinen, Sipi Kylmälä, and Aapo Peltola. 2018. “Binocular Disparity Can Augment the Capacity of Vision without Affecting Subjective Experience of Depth.” *Scientific Reports* 8(1):15798. doi: 10.1038/s41598-018-34137-9.

Rogers, Brian, and Maureen Graham. 1979. “Motion Parallax as an Independent Cue for Depth Perception.” *Perception* 8(2):125–34. doi: 10.1068/p080125.

Sachse, Pierre, Ursula Beermann, Markus Martini, Thomas Maran, Markus Domeier, and Marco R. Furtner. 2017. “‘The World Is Upside down’ – The Innsbruck Goggle Experiments of Theodor Erismann (1883–1961) and Ivo Kohler (1915–1985).” *Cortex* 92:222–32. doi: 10.1016/j.cortex.2017.04.014.

Sato, Makoto, Takumi Suzuki, and Yasuhiro Nakai. 2013. “Waves of Differentiation in the Fly Visual System.” *Developmental Biology* 380(1):1–11. doi: 10.1016/j.ydbio.2013.04.007.

Schall, JD. 1995. “Topography of Visual Cortex Connections with Frontal Eye Field in Macaque: Convergence and Segregation of Processing Streams | Journal of Neuroscience.” Retrieved January 6, 2022 (https://www.jneurosci.org/content/15/6/4464).

Soares, Sandra C., Rafael S. Maior, Lynne A. Isbell, Carlos Tomaz, and Hisao Nishijo. 2017. “Fast Detector/First Responder: Interactions between the Superior Colliculus-Pulvinar Pathway and Stimuli Relevant to Primates.” *Frontiers in Neuroscience* 11. doi: 10.3389/fnins.2017.00067.

Spering, Miriam, and Marisa Carrasco. 2015. “Acting without Seeing: Eye Movements Reveal Visual Processing without Awareness.” *Trends in Neurosciences* 38(4):247–58. doi: 10.1016/j.tins.2015.02.002.

Stensola, Hanne, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I. Moser. 2012. “The Entorhinal Grid Map Is Discretized.” *Nature* 492(7427):72–78. doi: 10.1038/nature11649.

Suri, Ikaasa, Patrick McGranor Wilson, Saba Doustmohammadi, Anna De Schutter, Thida Sandy Chunwatanapong, Juanyi Tan, Sara Divija Varadharajulu, Nicholas Hunter O’Connell, Archibald Lai, Sakshi Dureja, River Jonathan Phoenix Govin, Katsushi Arisaka, and Elizabeth Anne Falcone Mills. 2020. “Visual Perception in the Periphery: The Role of Covert Attention Vectors in the Extraction of Semantic Information.” 2020.08.02.231803.

Ta, Brian, Maria E. M. M. Silva, Kelly Bartlett, Umaima Afifa, Annie Agazaryan, Ricardo Canela, Javier Carmona, Emmanuel John L. De Leon, Alyssa Drost, Diego Espino, Guadalupe Espinoza, Kyleigh Follis, Paul Gan, Lauren Ho, Christina Honore, Emily Huang, Luis Ibarra, Tessa Jackson, Mira Khosla, Caominh Le, Victor Li, Trevor McCarthy, Elizabeth Mills, Sukanya Mohapatra, Yuuki Morishige, Nancy Nguyen, Ziyan Peng, Kimya Peyvan, Michael Phipps, Isabella Poschl, Jagannathan Rangarajan, Charysa Santos, Leonard Schummer, Sky Shi, Natalie Smale, April Smith, Divya Sood, Cindy Ta, Anna Tran, Michelle Tran, Rui Wang, Patrick Wilson, Nicole L. Yang, Megan Yu, Selena Yu, Aaron P. Blaisdell, and Katsushi Arisaka. 2022. “Visual Perception of 3D Space and Shape in Time - Part III: 2D Shape Recognition by Log-Scaling.” 2022.03.01.482004.

Uka, Takanori, and Gregory C. DeAngelis. 2004. “Contribution of Area MT to Stereoscopic Depth Perception: Choice-Related Response Modulations Reflect Task Strategy.” *Neuron* 42(2):297–310. doi: 10.1016/S0896-6273(04)00186-2.

Uka, Takanori, and Gregory C. DeAngelis. 2006. “Linking Neural Representation to Function in Stereoscopic Depth Perception: Roles of the Middle Temporal Area in Coarse versus Fine Disparity Discrimination.” *Journal of Neuroscience* 26(25):6791–6802. doi: 10.1523/JNEUROSCI.5435-05.2006.

Vanduffel, Wim, Qi Zhu, and Guy A. Orban. 2014. “Monkey Cortex through FMRI Glasses.” *Neuron* 83(3):533–50. doi: 10.1016/j.neuron.2014.07.015.

Vinck, Martin, Bruss Lima, Thilo Womelsdorf, Robert Oostenveld, Wolf Singer, Sergio Neuenschwander, and Pascal Fries. 2010. “Gamma-Phase Shifting in Awake Monkey Visual Cortex.” *Journal of Neuroscience* 30(4):1250–57. doi: 10.1523/JNEUROSCI.1623-09.2010.

Vinck, Martin, Thilo Womelsdorf, and Pascal Fries. 2013. “Gamma-Band Synchronization and Information Transmission.” Pp. 449–70 in *Principles of Neural Coding*. CRC Press.

Vishwanath, Dhanraj. 2014. “Toward a New Theory of Stereopsis.” *Psychological Review* 121(2):151–78. doi: 10.1037/a0035233.

Wexler, Mark, and Jeroen J. A. van Boxtel. 2005. “Depth Perception by the Active Observer.” *Trends in Cognitive Sciences* 9(9):431–38. doi: 10.1016/j.tics.2005.06.018.

Yacoub, E., N. Harel, and K. Ugurbil. 2008. “High-Field FMRI Unveils Orientation Columns in Humans.” *Proceedings of the National Academy of Sciences* 105(30):10607–12. doi: 10.1073/pnas.0804110105.