We Image with our Brain Not our Eyes

(The Necessity of integrating the Requirements of the Visual Cortex into AR/VR/MR(XR) Optic Systems)

Lost in Translation

Today's AR/VR/MR(XR) head-mounted Near to Eye (NTE) optic systems are presenting visual information to the brain of the user in a manner that does not conform with the “learned” process of human vision and thus are not fulfilling the requirements of the visual cortex that are needed for the user’s brain to accurately and comfortably interpret what they are seeing. This is not a software problem, it **is** an optics problem.

The deficiency of the visual information being provided to the brain by today's optic systems is the root cause of the multiple issues holding back this industry:

XR Device Capabilities
XR Device Costs
XR Software Development
User Experience
Enterprise Adoption
Consumer Adoption
And more…

The combination of all of these challenges prevents XR technology from reaching its potential. **Until an optic system delivers visual information to the brain as comfortably and natural as everyday sight, these challenges will persist and continue to impede the progress and scalability of XR tech.**

This paper explains briefly the “learned” process of human vision, the reasons why the current Near to Eye (NTE) optic systems are not meeting the requirements of the human visual process, the case for the absolute necessity of the optic systems to align with this process, and describe a simple solution that we will all reflect on for decades to come.

Device Demands

Near-To-Eye (NTE) optic systems come in many forms and utilize several different optical techniques to produce the imagery they generate. These devices are called Smartglasses, Head Mounted Displays (HMD’s), AR Glasses and VR Headsets, (and others). For this paper I will just use the term HMD to represent these devices.

If the use of the HMD is to view relatively simple imagery like text or info-graphics, the optical technique is not that critical. The brain can look at the words being projected or a graphic like an arrow, and understand what it is seeing. In that use case example you are just observing data and there is no need for correlating the imagery to the real world.

However, if you are trying to use an HMD to communicate real imagery and create an immersive, emotionally connected experience, like we have in the real world, the optical technique used to generate the imagery is critical. The optical technique is no longer just trying to put an image out in space in front of you, it is trying to either present a virtual world or augment the world. In this use case it is a fundamental necessity that the generated imagery is communicated with the brain and visual cortex in the same way we have learned to see.

This is an extremely complex issue that has been overlooked and/or not properly addressed in NTE imaging systems.

Learning to See

When human beings are born they are born with the ability to reach remarkable levels of physical and mental capabilities. When it comes to mobility we first learn how to crawl, walk, run, then jump and so on, perfecting our ability to move freely about. Some people take this development process to extraordinary levels and become Olympic or professional athletes.

This "learning process" happens with visual and mental development as well; we first perceive the world around us, then we identify objects and sounds, then we learn how to form concepts, then speech and so on learning to read and write. Some people perfect their mental acuity to such levels of refinement that they become scientists, mathematicians, physicists, economists, philosophers, chemists, doctors, engineers, lawyers, machinists, teachers and so on. The human potential is truly remarkable.

The process of learning how to see and to properly process visual data follows this same method of development. It takes a human more than 7 years to fully develop their visual processing capability [1]. There is some more recent scientific data that shows the visual system is still developing into the mid teen years [2]. Visual development is the most brain intensive process we have as humans. In fact more than 60% of the brain is used to process our vision system on a daily basis [2].

Each of these different areas of development; walking, talking, reading, thinking, and seeing have their own rules and processes for humans to become proficient at them. In order to work properly these rules must be followed and obeyed. For example the process of walking is actually a controlled process of falling forward, which requires leaning forward, but not too far. It takes a child a year or two to just master the basics of walking. The ability to read follows its own set of rules; the identification of symbols (the alphabet) to represent objects and concepts, punctuation, grammar, and even the spacing of the symbology. If you do not follow these rules it becomes almost impossible to understand what is trying to be conveyed. Here's a simple example of what happens when we try to read and understand a sentence without following the learned rules. *properinterpretationofsightistheresultofthelearnedprocessofintegratingobserveddifferentials*.

Following the learned rules of reading and writing properly, this sentence reads: *Proper interpretation of sight is the result of the learned process of integrating observed differentials*. This same principle of the necessity of following the rules applies to vision as well. If a Near-To-Eye optic does not deliver the visual information in the proper manner and form that the visual cortex is expecting it to be in (as it learned how to see), it cannot properly interpret the visual data.

Learning how to see is a very complex process. Depth perception, which is the ability to judge if objects are nearer or farther away than other objects, is not present at birth. It is not until around the fifth month that the eyes are capable of working together to form a three-dimensional view of the world and begin to see in depth.

A human's ability to focus their eyes, move them accurately, and use them together as a team must be learned. Also, babies and young children need to learn how to use the visual information the eyes send to their brain in order to understand the world around them, and interact with it appropriately. This process of learning how to see utilizes an enormous data base gathered for years to make sense of the light collected by the eyes.

The Visual System

The major components of the visual system can be broken up into visual acuity, depth perception, color perception, tracking, object recognition, sensitivity, and light sensitivity. The light data collected by the eyes is often referred to as mono and stereo cues. Some of these cues consist of the following:

contrast, hue, color, time of arrival of the light wavefronts, motion, parallax, shadows, texture, relative size, brightness, focus, occlusion, to name a few. In addition to these we also use physical techniques like head tilt, head rotation, eye quiver (or saccades), eye movement, blinking, and squinting to assist in our perception of the things we see.

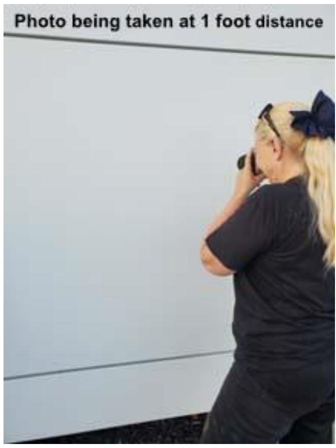
If the light information presented to the eyes is in the form that we learned to see, the world makes visual sense. However, just like the reading example above, if the rules are not followed properly, it can cause visual confusion as well as undo stress on the eyes and brain.

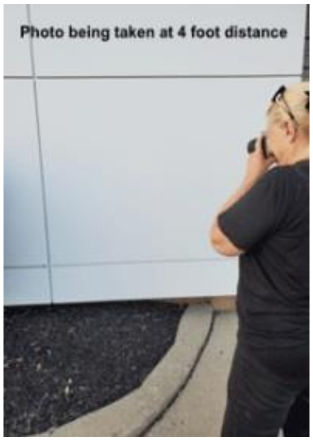
Context Requirements

Below is an example of 4 pictures taken of a wall from four different distances. Because there is not enough visual data in the light entering the eye, there are insufficient mono and stereo cues for your depth perception to work properly. It is impossible to determine the distance of the wall.

These photos demonstrate that simply changing the focal plane distance of an image does nothing to help the brain understand depth. Humans determine depth and see 3 dimensional objects in a volume, this is a learned process and consists of processing light differentials. If there are not enough differentials, your brain does not know how to interpret what it sees.

The following series of photographs of a monochromatic wall are taken at different distances ranging from 1 foot to 8 feet. You will notice that it is impossible to determine the distance of the wall because there is not enough data for your brain to properly process the imagery to provide an answer. The camera was focused to the proper distance for each photo. Even so, adjusting the image plane distance for the camera to focus properly does not provide the required information for your visual cortex to understand depth.





Attempts to Solve

There are some companies and researchers who have postulated that changing the focal distance of the image plane in an HMD will help in determining depth of an object. As you can see from the images above this is a serious over simplification of what is actually required to determine depth. This same postulation incorrectly asserts that varying the image plane will solve the Vergence/Accommodation conflict [VAC] [3]. (see https://www.cs.umd.edu/sites/default/files/scholarly_papers/Kramidarev.pdf).

Using the method of variable focal distance to assist with depth perception in an HMD has very little to do with how humans determine depth. It **is** how cameras determine depth – not humans.

Interpretation Issues

The problem of properly interpreting imagery is not limited to simple, monochromatic imagery. The following examples contain more complex imagery but still do not provide enough data to fully understand the images.

We have all seen photographic optical illusions, often involving fish. In each of the images below it is difficult to determine the actual size of the fish. Is this sturgeon 24 inches long and much closer to the camera or is it much bigger, more like 10 feet long.

How big is the Sturgeon?

How big is this shark?

Or this shark?

How big is this Grouper?

[the Sturgeon is 5 inches long, the sharks are both 7 inches long, the Grouper is 10 feet long]

In order to know the actual sizes of these fish requires more data than these images provide. In the real world if each of these fish were in your presence you would know how big they are because your eyes would be collecting sufficient data that is entering your eye naturally and unadulterated.

Let There Be Light

The critical characteristic of natural light is that it comes from many different angles and distances and contains an incredible amount of information. This type of light as it enters the eye is called a freeform wavefront. Freeform wavefronts carry their own unique time signature. This is their primary distinction from planar wavefronts. Freeform light carries all of the essential mono and stereo cues that our visual cortex and brain need to properly process and interpret what you are seeing. This is how we learn how to see.

The critical point for HMD NTE optic design is that the visual cortex has “learned” very specific techniques and processes of how to see, and that these same rules must be followed in delivering the imagery (light) into the eye from NTE systems, to eliminate confusion and reduce eye strain.

It is typically stated that humans see the world in 3D; it is far more accurate to say that we see 3 dimensional objects placed in the world volumetrically. As light comes towards us from many angles, the eye collects the light data and the optic nerve converts the data into electrical signals which are then processed in the visual cortex of the brain. This electrical information is then rendered into an image.

It’s About Time

The actual perceived volumetric imagery of vision comes as a result of each piece of light reaching the eye at slightly different times from many angles. It is the time differential that creates the volume we see. This fundamental point cannot be overstated.

Time is **the** major contributing factor used by the brain’s visual cortex to determine depth coupled with the mono and stereo cues. Stereoscopy or parallax is but one component of depth perception. If stereoscopy were the main factor resulting in our volumetric perception of the world (often called a 3D view), everything would collapse into a flat plane if we covered one eye which, of course, it doesn’t.

It is the multi-angular freeform wavefronts and their time differential that creates the 3 dimensional volume we perceive. These freeform light wavefronts provide the rich data the brain and visual cortex need to determine depth properly.

We see What We’ve Learned to See Regardless

The visual cortex’s learned processes of sight are so powerful and engrained in us that they sometimes override what we know to be true. If you have never seen a demonstration of how powerful the brains visual cortex is please take 2 minutes to view this Charlie Chaplin Demo.

The demo shows a hollow face mask that you could wear, rotating on a turntable. This is an image looking at the Charlie Chaplin face mask from the front of the mask; it is a convex face, as all faces are convex.

Here is an image of the same facemask from the backside, which is concave but appears to be convex. Your brain is telling you the nose is pointing towards you, but it's not, it is pointing away. Since the first moment you opened your eyes as a baby you have never seen a concave face, and therefore your brain sees what it thinks you should see. Please watch the 2 minute video, it is fascinating. Here is the link: https://youtu.be/QbKw0_v2clo

Here is an additional link to 10 optical Illusions. These are amazing and demonstrate some of the extraordinary attributes of the human visual system. Several of these illusions defy explanation, and simply contribute to the point that it is imperative that we integrate "how" we see into the design of NTE imaging systems.

https://www.youtube.com/watch?v=_t2ePlwTeBQ

We See with our Brains not our Eyes

Much of the effort over the last 20 years to design a NTE optical system has centered on how the eye functions, not the entire visual system. These efforts have emulated how the eye directs light to the retina, which was the same process used to develop the camera. The primary deficiency of current optical methods of NTE imaging systems is that they are based on the principles of camera optics. Lens based optical systems compress time to a single plane, the image plane. This is how we define and focus a camera image to the film or sensor. There is no volume to a photograph because the optics have compressed time to a single value. The brain never does this.

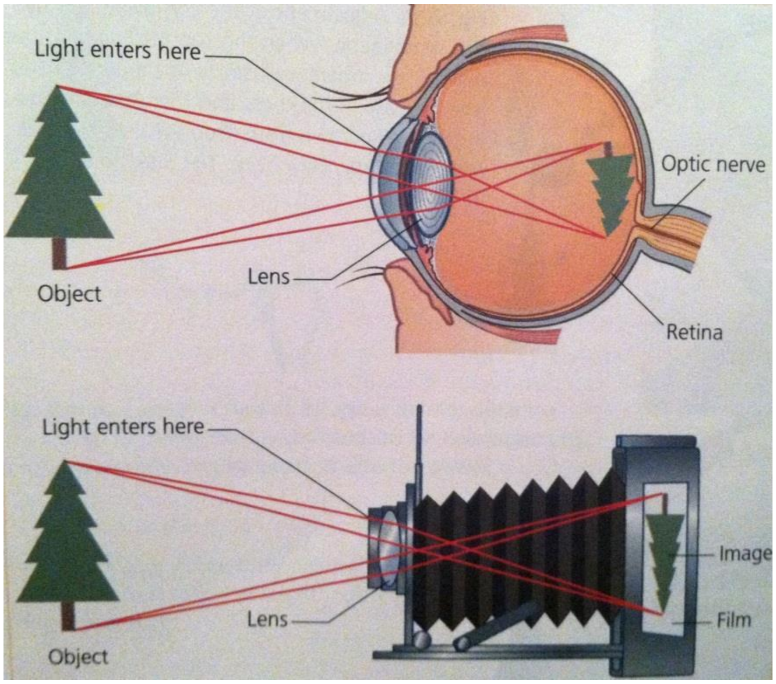
The individual components of the eye work in a manner similar to a camera, but with a couple critical differences. The similarities start with the cornea which takes widely diverging rays of light and bends them through the pupil, the dark, round opening in the center of the colored iris. The iris and pupil act like the aperture of a camera. The light then goes through the lens of the eye, which acts like the lens in a camera, helping to focus light to the back of the eye, the retina.

The retina is similar to the film in a camera. The retina is covered with nerve cells called rods and cones, which convert the light rays into electrical signals and sends them through the optic nerve to the brain where an image is perceived.

Our Eyes are Windows not Cameras

The critical differences between a camera and the eye is the eye's ability to rapidly look at near and far objects and focus on them at will, combined with the eyes ability to collect light from many angles and distances simultaneously and project them on a compound curved surface (the back of the eye). The eye does not compress time, the curvature of the retina preserves the volumetric time signature of the incoming light. Lenses compress the time signature to a single value, which is the flat image plane of the camera. The combination of the diverse angular collection combined with the different time signature of each piece of light produces what we call a "light field", or a "volumetric" view.

Here are images of the camera compared to the eye:



And a schematic of the incredible processes in the brain to produce an image:

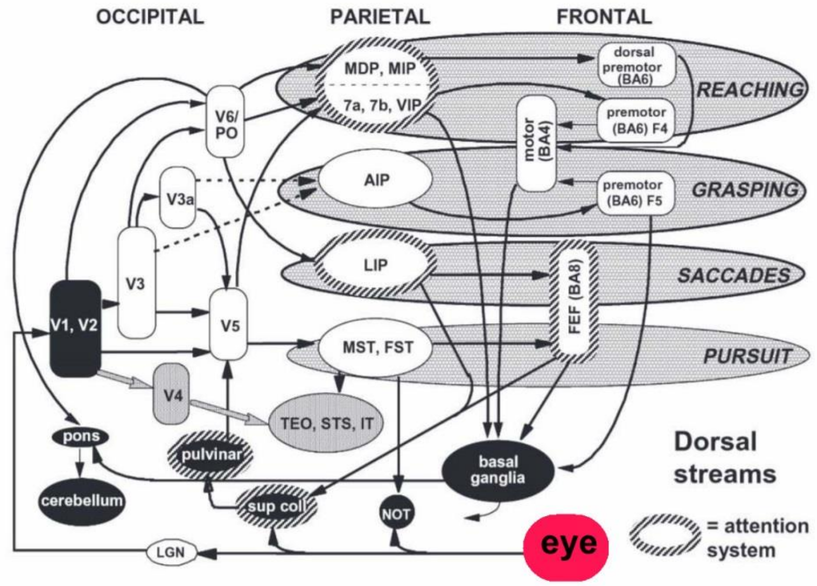


Fig. 1. Brain areas involved in visuo-motor modules and their connections, for four aspects of the development of visually controlled behaviour. Areas involved in spatial attention networks are highlighted by hatched borders. Based on data reviewed by Jeannerod (1997), Milner and Goodale (1995) and Rizzolatti et al. (1997). (Redrawn and updated from Atkinson and Braddick (2000).).

Eyes Collect Light and Brains Produce Imagery

The complex process of creating an image from the electrical signals sent from the retina to the brain is only partially understood. You can see that the function of the eye is that of a light collector. It is the visual cortex and the brain that produce the imagery. This is why it is so critical to integrate the entire visual system into NTE optical systems. The visual data must be delivered to the visual cortex in the same manner in which we “learned” to see in order to be able to make proper sense of the data.

Wavefronts: Planar vs. Freeform

The main difference between current optical NTE optical systems and how we see in the real world is that HMD's produce an image plane and deliver a planar light wavefront into the eye. There are no planar wavefronts in the real world, an image plane is a man-made contrivance. We did not "learn" to see the world with planar wavefronts, we "learned" to see with freeform wavefronts.

When NTE optical systems deliver planar wavefronts into the eye there are insufficient mono and stereo cues, with correct timing, to properly process and interpret the visual data. This causes the vergence/accommodation conflict and eye strain. There is not enough information for the brain to properly determine depth. Moving the image plane back or forth does not correct, or provide the missing data. It is not possible to force the perceived distance of an object. Proper distance and correlation to the real world can only be accomplished by providing the light data in the way we "learned" to see and allowing the visual cortex and brain to properly process the data.

Flat vs Curve

One relatively simple method of delivering freeform light wavefronts into the eye is to use multiple freeform, off-axis compound curved, first surface mirrors. This technique helps to reinstate the original time signature of the captured imagery. The current limitations of display technologies do not allow the full effects of this technique to be realized. That subject is beyond the scope of this current paper and will be addressed separately. The curved first surface mirror method does not cause the vergence/accommodation conflict in the first place so there is nothing to correct.

The curved mirrors create multiple viewing angles of the source imagery, allowing the eye to move around freely, absorbing freeform wavefront light energy from many different angles. This natural process of absorbing and processing both mono and stereo cues produces an in-focus image at any distance by providing sufficient data, including the time differentials for the visual cortex to properly interpret the imagery.

Give Them What They Need

Working in harmony with how the Human vision system "learned" to see should be the central theme of HMD NTE optical design. AR and VR headsets must present freeform wavefront imagery as natural and unadulterated as possible into the eye, just like the real world. The visual cortex and the brain will do the rest. Compound curved, first surface mirror optical systems address the basic need for freeform wavefronts. This method produces a Natural Eye Optic which facilitates Human comfort, Human safety, Eye comfort, and Eye safety and provides a great user experience.

Douglas Magyari

CTO

IMMY Inc.

Doug Magyari CEO/CTO IMMY Inc.: Inventor, scientist, entrepreneur, and visionary, Doug has spent more than 20 years working with immersive technologies and developing augmented reality (AR) headsets. He has 15 patents in chemistry, optics, acoustics, mechanics, and electronics with additional patents pending. He has started, operated, and sold six different companies.

For more information on Volume Imaging, the IMMY Mirror Optic System, and the IMMY Immersive Imaging Glasses visit IMMYinc.com or contact us at: info@immyinc.com

References:

1. **Susan J. Leat, Naveen K. Yadav, and Elizabeth L. Irving, (2009) "Development of Visual Acuity and Contrast Sensitivity in Children" Journal of Optometry Volume 2, Issue 1 , 2009 , Pages 19-26** https://www.sciencedirect.com/science/article/pii/S1888429609700196

2. **Oliver Braddick, Janette Atkinson, (2011) "Development of Human Visual Function" Vision Research Volume 51, issue 13, pages 1588-1609** https://www.sciencedirect.com/science/article/pii/S004269891100068X

3. **Gregory Kramida and Amitabh Varshney (2015) *IEEE Explore,* IEEE.org "Resolving the Vergence_Accommodation Conflict in Head Mounted Displays".** https://ieeexplore.ieee.org/abstract/document/7226865/authors#authors

Author Keywords

Augmented Reality; Virtual Reality; Mixed Reality; Vergence Accommodation Conflict; Volumetric Imagery; Lightfield.