Most Effective Capabilities of Head Mounted Displays for Dismounted Soldier Training Using Augmented Reality

John Baker, Scott Johnson,
Jaime Cisneros, and Juan Castillo
Chosen Realities LLC
Orlando, FL
[john.baker, scott.johnson, jaime.cisneros,
juan.castillo]@chosenrealities.com

Pat Garrity
U.S. Army Research Laboratory
Orlando, FL
patrick.j.garrity4.civ@mail.mil

ABSTRACT

Within training, the Department of Defense (DoD) has a strong interest in augmented reality (AR) for its ability to combine live and virtual assets to reduce cost, increase safety, and to mitigate unavailability of needed live assets. During the past two years, a rapid increase of interest in AR for consumer use has spawned a multitude of innovations for head mounted displays (HMD). Increased fields of view (FOV), tetherless computing, integrated depth-sensing, external spatial audio, and simultaneous location and mapping (SLAM) are just a few features that have become a boon for military use, such as dismounted Soldier training. However, the usefulness of these features varies for the dismounted Soldier training use case. This paper examines features from nearly a dozen of today's consumer AR HMDs and contrasts the tradeoffs required to reap their benefits. We evaluated these HMDs primarily against key tactics and skills required in ATP 3-21.8 'The Infantry Rifle Platoon and Squad' doctrinal framework – specifically employing fires, offensive operations, defensive operations, and patrolling. Finally, this paper explores what features are missing or are suboptimal on these HMDs for dismounted Soldier training use.

John Baker is the Managing Director and Chief Operating Officer for Chosen Realities LLC. He shared the 2012 Modelling and Simulation Award from the Army for his work as Chief Engineer on the Dismounted Soldier Training System (DSTS). John has over 20 years of experience in defense, building a wide variety of systems in various capacities. He has a Bachelors in Industrial Engineering from the University of Central Florida and a Masters in Industrial Engineering from the Pennsylvania State University.

Scott Johnson is the Chief Technology Officer at Chosen Realities LLC. He shared the 2012 Modelling and Simulation Award from the Army for his work as Lead Software Engineer for (DSTS). Scott has 12 years of experience in defense and 10 years of experience in the video game industry where he worked primarily in the areas of Animation and Physics. He has a Bachelors in Electrical Engineering from Purdue University and a Masters in Computer Science and Engineering from the University of Michigan

Jaime Cisneros is a Principal Engineer at Chosen Realities LLC. He has 24 years of experience in constructive and virtual simulation and training. He was the simulation lead for the Dismounted Soldier Training System (DSTS). As a researcher, Mr. Cisneros has designed and developed advance prototypes for a next generation virtual reality dismounted soldier trainer and has published nearly a dozen papers in this field. He has a Bachelors and Masters in Computer Science from the University of Central Florida.

Juan Castillo is a multi-discipline engineer and 3D artist at Chosen Realities LLC. He has 8 years of experience in virtual reality, gaming, and graphics industry. He has a Bachelors in Computer Animation from Full Sail University.

Pat Garrity is a Chief Engineer at the U.S. Army Research Laboratory-Human Research and Engineering Directorate, Advanced Training and Simulation Division (ARL-HRED-ATSD). He works in Dismounted Soldier Simulation Technologies conducting research and development in the area of dismounted soldier training and simulation where he is the Army's Science and Technology Manager for the Augmented Reality for Training Science and Technology Objective (STO). His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He has his B.S. in Computer Engineering from the University of South Florida and his M.S. in Simulation Systems from the University of Central Florida.

Most Effective Capabilities of Head Mounted Displays for Dismounted Soldier Training Using Augmented Reality

John Baker, Scott Johnson,
Jaime Cisneros, and Juan Castillo
Chosen Realities LLC
Orlando, FL
[john.baker, scott.johnson, jaime.cisneros,
juan.castillo]@chosenrealities.com

Pat Garrity
U.S. Army Research Laboratory
Orlando, FL
patrick.j.garrity4.civ@mail.mil

INTRODUCTION

Augmented reality (AR) is valuable to dismounted Soldier training because it removes many of the limitations and costs associated with live training. For example, live training makes use of live role-players for OPFOR and civilians, which are limited in number because of the high recurring costs. As a consequence, the limited number of these role-players reduces the sense of urban sprawl that a real urban area might have. AR can create an unlimited number of virtual OPFOR and civilians – all appearing embedded realistically within the training area, going about their daily routine amongst live Soldiers. Furthermore, AR can provide capabilities beyond live training such as making the daytime appear as nighttime on-the-fly, creating vivid backdrops of entire cities (e.g., make Ft. Irwin appear as if it were within the middle of Baghdad), and placing virtual battlefield effects as close to Soldiers as needed without worry of safety concerns. Head mounted displays (HMDs) are a key component of AR.

This paper evaluates various commercial off-the-shelf AR HMDs against various critical capabilities for dismounted Soldier training. There are many AR HMDs to consider, with all vendors claiming to have “the best”. This paper’s intent is to help the reader consider an HMD based on important characteristics that are typically not cited as part of an HMD’s public specification – or even known at all by their respective vendor. In any case, although the industry is rapidly maturing, it is our position that current AR HMDs have some key gaps that still need addressing.

For the purposes of this research, we consider “mixed reality” as commensurate with AR, since both have the identical goal of inserting virtual assets into a live scene. We also classify HMDs into two different groups:

1. Optical See-Through. These HMDs project virtual object imagery into and/or onto a transparent medium that directs this imagery to the user’s eyes. Because this HMD uses a transparent medium, the user can see the live scene naturally – and the virtual assets appear inserted into the live scene. The Microsoft HoloLens is an example of an optical see-through HMD.

2. Pass-Through Video. These HMDs use a video camera(s) that streams the entire live scene onto an opaque medium (e.g., a smartphone-like screen). Virtual asset imagery is combined with this video, making the imagery appear inserted into the live scene. The VRVana Totem is an example of a pass-through video HMD.

We evaluated eleven different AR HMDs – all of which were classified by their respective vendors as being a production unit, a development kit, or a prototype representative of what they intend to release publicly. These HMDs were the latest products by select vendors as of June 1, 2017. A subset of the HMD vendors gave us permission to explicitly use their product’s name in this paper; whereas there were others that did not allow any attributable disclosure. At the risk of unintentionally disclosing non-public HMD data through process of elimination, no HMD product names will be used henceforth. The goal of this paper is not to guide the reader into which HMD should be purchased, but to instead guide the reader into considering various characteristics of HMDs given specific use cases.

The HMDs evaluated had an average price of $6,875 and a median $3,000. They had an average diagonal FOV of 66.45 degrees and a median of 41 degrees. For dismounted Soldier training purposes, the HMD must be less than two pounds (“Two New Soldier Optics,” 2015; “Enhanced Night Vision,” 2017) on the Soldier’s head – of which, all of the evaluated HMDs qualified.

BACKGROUND

As for dismounted Soldiers training: It is the training of Soldiers that are not in or on a vehicle. The US Army makes use of 'ATP 3-21.8 Infantry Platoon and Squad' as a primer in dismounted Soldier training curriculum. Unlike vehicle or platform training, dismounted Soldier training is unique in that the Soldier carries training equipment on his/her body while needing to have free movement in constrained spaces. Soldiers "train as they fight" – and this presents the further challenge of minimizes any encumbrances to the Soldier in dismounted training.

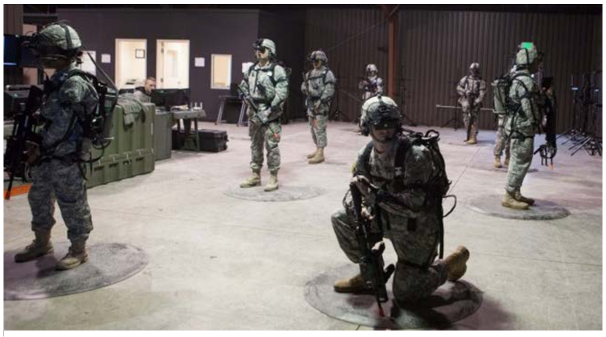

Figure 1. Dismounted Soldier Training System (DSTS) in use. (Zamora, 2013)

We can draw parallels and lessons learned from comparing historical training system. For example, the Dismounted Soldier Training System (DSTS) is a virtual reality-based system the Army uses for dismounted training purposes. It can be used as a starting point as to what systems seeking to use similar technology should and should not embrace – especially AR-based systems. Since this paper focuses on HMDs, it is valuable to use some lessons learned from DSTS's HMD. Here are three key takeaways:

- The HMDs used in DSTS too frequently hampered the Soldier's ability to have an acceptable cheek weld on his/her weapon – a key element of "train as you fight"; if the HMD's form factor were slightly smaller, this would not have been an issue.
- DSTS has a motion tracker affixed to the top of the Soldier's helmet because the HMD does not have its own integrated tracking capability. This presents three problems. First, because the motion tracker is not integrated into the HMD, it had to have its pose calibrated against the HMD at the start of every exercise – this adds complexity and setup time. Second, for a variety of requirements, this motion tracker used a wired connection to other parts on the Soldier – this adds encumbrance to the Soldier because there is a chance this wire would snag on the Soldier's gear. Third, the tracker calculates yaw using exceptionally small gyroscopes, which is a very common method. Over time, these gyroscopes accumulate error (LaVelle Yershova, Katsev, & Antonov, 2014) insomuch that its orientation will no longer match that of the HMD. The most common technique to correct this error is to use the earth's magnetic field as a stable reference point. However, the earth's magnetic field is not reliably stable in environments that have ferrous materials (e.g., steel I-beams in walls, metal piping underground, rebar in the foundation). In short, yaw calculation using this method was not reliable across several Army posts found worldwide.
- The HMD's field of view is 60 degrees horizontally and 48 degrees vertically – about 77 degrees diagonally, which was world class when DSTS fielded in early 2012. However, the field of view was still too small. We participated in an extensive study sponsored by the Joint Non-Lethal Weapons Program in the summer of 2015. This study had participants from the Army, Marines, Air Force, Navy, and DHS performing various scenarios with DSTS. Participants stated that the "narrow FOV and the consequential lack of peripheral vision are cited as problematic" and that the FOV was "inadequate for proper situational awareness."

Both VR and AR are commonly used in the same contexts (e.g., gaming, training) and, as such, seek to provide the same result: Realism. However, the differences in demands between VR and AR are wide, so using DSTS lessons learned has some limitations. The table below shows a non-exhaustive list of common features and a sense of the relative difficulty in providing each. It is our opinion that there are far more differences between VR and AR than there are similarities. Both VR and AR will display virtual objects typically via an HMD, but this is where their similarities end.

Table 1. Major differences between VR and AR

Virtual Reality	Augmented Reality
Dynamic occlusion (i.e., moving objects occluding other objects) is handled by game engine. This is simple.	Dynamic occlusion is very difficult, requiring high performance and high fidelity depth-sensing.
Audio automatically comes from the game engine. This is relatively simple.	Audio comes from two sources – the game engine and live environment. Mixing is non-trivial.
All objects can appear with the same fidelity.	The fidelity between live and virtual objects greatly vary (so far). For optical see-through HMDs, the live objects have higher fidelity than virtual objects. For pass-through video HMDs, the live objects usually have lower fidelity than virtual objects; this is because the input video cameras have a lower resolution than the screen that's being used as the display medium.
Objects always appear opaque because these HMDs universally use physically opaque display mediums (e.g., an OLED screen).	Object will only appear opaque with pass-through video HMDs. There is one optical see-through HMD that appears to have solved this but it is still in prototype at the time of this writing. We also note that the use of photochromic lensing on optical see-through HMDs helps, but it is not truly opaque.
Changes in ambient light won't affect visuals because the HMDs' functionality is independent of ambient light.	Changes in ambient light will always change visuals for any type of HMD. Optical see-through HMDs' virtual objects will change in transparency. Pass-through video HMDs' video cameras can lag with light changes.
Shading and shadowing of objects is handled by the game engine, which is simple.	Shading and shadowing of objects – both live and virtual – is very difficult (e.g., applying virtual light onto live objects). Optical see-through HMDs are additive in display (i.e., no black pixels), which means shading lacks realism.

EVALUATION

We've selected four major sets of tactics and skills areas in ATP 3-21.8 that represent the widest range of uses for HMDs for dismounted Soldier training: employing fires, offensive operations, defensive operations, and patrolling. These were selected based on the demands for the Soldier to move freely in spaces of widely varying size across an unrestricted distance – such as a Military Operations on Urbanized Terrain (MOUT), where Soldiers will train in and around simulated towns with varying terrain. See figure to the right.

We selected a *non-exhaustive* list of ten characteristics of HMDs that are needed to facilitate these four tactics and skill areas. The HMDs will be evaluated against these characteristics in the subsequent sub-sections.

Figure 2. Muscatatcuk MOUT, Indiana. ("Muscatatuck Urban Training Center", n.d.)

- Cheek weld
- Dynamic occlusion
- Situational awareness
- Visual acuity
- Use of scopes
- Pose
- Tracking body parts
- Mixing live/virtual audio
- Shading and shadowing
- Physical encumbrance

Cheek Weld

According to 'FM 3-22.9 Rifle Marksmanship M16-/M4 Series Weapons', the cheek weld, also called cheek-to-stock weld, will "…provide a natural line of sight through the center of the rear sight aperture to the front sightpost and onto the target. The firer's neck should be relaxed, allowing his cheek to fall naturally onto the stock." This is practiced until the cheek weld is applied quickly with consistency. It is considered to be a critical element in firing one's weapons. Thus, the sides of the HMD cannot impede the Soldier's cheek from sitting on top of the butt stock.

Which HMDs are best capable of providing a good cheek weld? A cheek weld can be impeded because the HMD protrudes too far forward. It is doctrine ("FM 3-22.9," 2008) that Soldiers are taught to position their cheek welds so that their nose barely touches the charging handle. Soldiers are free to go further back, but this gives us the limiting constraint because a Soldier's head cannot go any further forward than this. Thus, the HMD would ideally not protrude any further forward than the protrusion distance of the Soldier's nose to his/her face. To calculate this distance, we use the length of adult noses and their respective angles in relation to their faces (i.e., nasofacial angle) to derive that this distance is a minimum of 1.95 centimeters within 95% confidence (Zankl, Eberle, Molinari, & Schinzel, 2002; Wen, Hai, Lin, Guosheng, & McGrath, 2015). This represents the maximum distance an HMD can protrude from where it sits on one's nose (specifically the nose root, which is as far up a nose one can get) to ensure that 95% of the population can cheek weld with their nose touching the charging handle.

Only two of the eleven HMDs evaluated were below 1.95 centimeters for this measurement, ensuring success for 95% of the population. If strict averages were used across the entire population, we get 2.93 centimeters – this increases our count to three of the eleven HMDs.

Dynamic Occlusion

Occlusion is the ability for live objects to partially or fully occlude virtual objects, which is considered to be critical for virtual objects to give the illusion that they are part of the world (Livingston, Brown, Swan, Goldiez, Baillot, & Schmidt, 2005). Live objects that are static, such as walls, have the convenience of being scanned before operational use. These scans provide occlusion information to virtual objects that may appear in an exercise; we refer to this as *static occlusion*. In contrast, *dynamic occlusion* is concerned with live objects that are moving – scanning before operational use is useless. The 3-dimensional profile of these live objects need to be translated into the virtual world quickly (i.e., high framerate), when movement occurs (i.e., low latency), and provide a realistic boundary between the live and virtual objects (i.e., high fidelity). This is limited by the range of the depth-sensing capabilities that drive dynamic occlusion.

We see dynamic occlusion as one the most challenging capabilities to bring to AR because providing high framerate, low latency, and high fidelity are rife with challenges that can extend beyond the small form factors expected from an HMD. The challenges begin with how depth information is captured, which is (so far) universally through one or more of these techniques: stereovision, time of flight (ToF), or structured light.

- Stereovision relies on the use of two cameras to identify and compare the same point(s). Based on how each camera sees the point(s) and the two cameras' inter-pupillary distance (IPD), the depth to the point(s) can be calculated. Stereovision generally has large minimum distance versus the other depth-sensing methods. This minimum distance is frequently greater than a foot, which makes capturing depth up close difficult, inaccurate, or impossible. This is germane to the dynamic occlusion caused by a Soldier raising his/her weapon into firing position, as the weapon will be very close to the HMD and well within the minimum distance so dynamic occlusion won't be possible. Additionally, stereovision relies on finding visible points to compare to calculate depth. Surfaces devoid of any features (e.g., a barren wall) won't provide depth information.
- ToF measures depth through measuring the roundtrip emission and detection of light – which is usually infrared (IR). This technique can be favored over stereovision mostly because it requires less form factor to house and its calculations generally produce less error. However, it suffers from interference from sunlight.

Sunlight has IR within its energy spectrum and this can confuse the ToF calculations. Furthermore, IR is absorbed by darker materials, rendering it unable to detect depth.

- Structured light is the use of emitting a set pattern of light, usually in IR. This pattern is then seen by on-board cameras. Distortions in the pattern from the scene translate into depth information. This has the same issues with the sunlight as ToF does.

All of these, however, have challenges for effective maximum distance. The further away objects are, the greater the error of the depth-sensing technology and this will cause the dynamic occlusion to look blocky or wavy. Consider the use of a stereovision camera, whose maximum distance is proportional to its IPD: The wider the IPD, the larger the form factor of the HMD and the greater the weight. As a warning to the reader, a stated maximum range for depth-sensing should be approached with caution: There are no industry standards on what defines an effective range.

Industry is actively working to solve all of the issues discussed above and they are expected to overcome these challenges in the coming years. Of the HMDs we evaluated, four of eleven had dynamic occlusion capability built-in. Another one was claiming to have this “soon” but at the time of this writing, we were unable to evaluate any early versions. The dynamic occlusion frame rate ranged between 0.25 Hz to 30 Hz. This is lower than the industry desired for displaying virtual objects, which is 90 Hz (Carmack, 2014; Binstock, 2015) as first set by Oculus.

As for the quality, this is subjective and we turn to how the green screen industry measures quality – since it seeks to “cut out” live objects and place them in an alternate environment (i.e., virtual object). We see this with weather forecasters on live television regularly, so we ask ourselves, “Do any of these HMDs possess the same quality?” The answer is ‘no’. Although green screening has a very controlled environment (e.g., lighting and distance), we would expect Soldiers to hold dynamic occlusion to the same standard. The most common problem seen with dynamic occlusion – other than slow frame rate – are large, blocky outlines that surround and encroach the live objects in view. When evaluating dynamic occlusion, we first recommend testing at short distances to ensure that a raised weapon can occlude its part of the scene. Also test amorphous objects, such as waving curtains, pouring water, and plants blowing in the wind – this is a great stress test for both framerate and fidelity.

Situational Awareness

Situational awareness (Endsley, 1995) is the ability to project future status given the perception of current environmental elements. The FOV is a significant contributor to this, and a narrower FOV has been shown to increase the time it takes to complete tasks (Ren, Goldschwendt, Chang, & Hollerer, 2016). However, there are two significant characteristics to consider.

The first characteristic deals specifically with the majority of optical see-through HMDs: The FOV that is able to display virtual objects is less than the FOV the Soldier can actually see in the live world. Contrast this with pass-through video HMDs in which these two FOVs are equal. When the virtual FOV is less than the live FOV, we call this less-than-live FOV (LTL-FOV). See figure on right. When a Soldier experiences LTL-FOV, they will expend some amount of cognitive energy to realize they must scan with their head to see all potential threats – not unlike what they do with night vision goggles (NVG) – so that their virtual FOV can cover their entire field of regard (FoR). The difference, however, is that this can produce negative training: Soldiers may find themselves scanning with their heads in combat when they shouldn’t. Furthermore, a wider FOV leads to improved performance for similar tasks, such as aerial door gunnery training (Stevens & Kincaid, 2014). Of the eleven HMDs we evaluated, seven had LTL-FOV, which were all of the see-through optical HMDs.

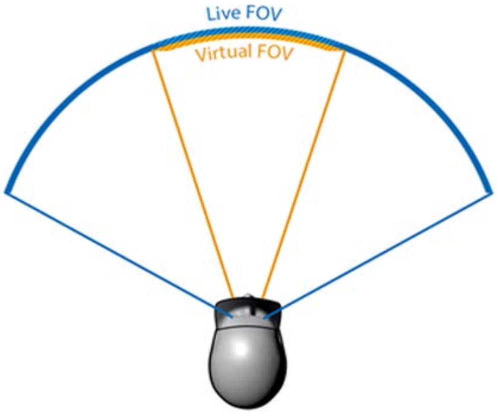

Figure 3. LTL-FOV

The second characteristic is specific to pass-through video HMDs and the properties of the video cameras they use. Video camera properties are deserving of their own research paper alone. However, we feel one of the most significant video camera characteristics to consider is the video pass-through HMD's ability to show luminosity naturally, that the luminosity in the view can show all *live* objects between light and dark areas. This is called *dynamic range*. The two images below show the exact same scene but with different exposure values, also known as *stops* or *f-stops*, which is based on the camera's aperture range. The image on the left, clearly shows the sun and the features of the clouds; whereas the image on the right has drowned out the sun and the clouds, but the people walking on the beach can now be seen. Imagine a Soldier not being able to see threats because of this.

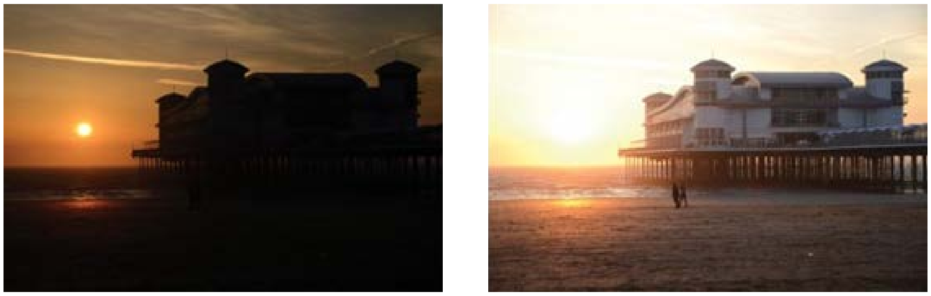

Figure 4. Example of dynamic range. Courtesy: TechRadar.com

Ideally, these HMDs need to match the dynamic range of human eyes, if possible, which is about 20 stops per frame (Vandenberghe, 2010). To date, there are not any video cameras that we can find that are small enough to fit in an HMD and/or perform this capability in real time. None of the vendors have disclosed how many stops their video cameras can support. Also, empirically measuring dynamic range requires that we are able to record the video feed from these HMDs to measure various values – and not all of the four video pass-through HMDs had the ability to record the raw video feed, which is required for the methods that we are familiar with (troy_s, 2016; Freriks, 2012). So this then becomes a subjective measurement based on observation. For all four of these HMDs, looking at scenes with high contrasts (e.g., a room with its lights off, but sunlight entering through a window) in lighting resulted in the inability to see all live objects clearly. Thus, pass-through video HMDs will currently require that training exercises be mindful of how much lighting contrast is present. We recommend that the reader perform their own tests when evaluating pass-through video HMDs to suit their needs.

Visual Acuity

Visual acuity, in the simplest sense, is the ability to see patterns or objects – both live and virtual – at a given distance relative to normal vision (National Institute for Standards and Technology, 2006). Aside from the issues that arise from luminosity, pass-through video HMDs may also show the live objects as being blurrier than virtual objects. This is largely due to the resolution of the video input cameras is spread too thinly across the field of view. For example, a 1920x1080 resolution video camera feeds a horizontal FOV of 100 degrees results in 19.2 pixels across each degree would result in about 1.1 pixels per 1.0 centimeters of an object that is 10 meters away. Research suggests that facial recognition requires about 1.25 to 5 pixels per centimeter (Axis Communications, 2014), depending on the condition, which would mean that this HMD would need to be 2.2 to 8.8 meters away from someone to recognize their face. The general standard for visual acuity is to have at least 60 pixels per degree (PPD) (Bailey, Wilz, & Arthur, 2012). With that, the 60 PPD is also held to virtual objects; however, virtual objects' resolutions are also limited by the GPU driving the graphics – the higher the resolution of the graphics, the slower the framerate.

There isn't a recommended minimum resolution for visual acuity because visual acuity is also a function of FOV. All of the HMDs evaluated had resolutions up to 1440p, but this is spread across their FOVs, resulting in an average PPD of 31.28 (median of 29.37). Only one scored above the recommended 60 PPD.

Use of Scopes

The capableness of an HMD working with a scope depends on what implementation of a scope will be used. The scope needs to be able to portray AR, mixing both live and virtual objects together realistically. There are two methods that are typically used. The first method physically recreates the same AR technology found in an HMD but in a monocular manner. The second method creates a virtual scope, which is overlaid on the live weapon – wherein any zoom is handled digitally by at least one of the HMD's video camera feeds, if it has any. Note with this second method, the weapon's position and orientation needs to be very precise in relation to the HMD so the virtual scope appears properly. Using a digital magnification in this manner is non-trivial and may not produce images with high enough resolution from any of the HMDs we evaluated. This is dependent upon the required magnification of the scope (e.g., an ACOG at 6X). Furthermore, this method may not be feasible on red dot optics such as the M68 CCO because the HMD's video feed camera may not have a wide enough FOV to provide an unaligned eye-to-weapon shot.

AN/PAS-13(M), AN/PAS-13(L), M68 CCO, ACOG, AN/PVS-4, and M145 MGO were the simulated scopes required for DSTS and we use them here in our evaluation as well as an adequate crossover requirement. These physical scopes can be moved forward enough on a weapon that proper eye relief could be achieved from any of the HMDs we evaluated – some HMDs protrude forward quite far. However, the limiting factor was not eye relief, rather, it was collision with the charging handle – and this topic has already been covered in the Cheek Weld section.

Moving the physical scopes forward will cause a change in the weight distribution of the weapon and Soldiers may notice this when they aim – the scopes (since they are simulated) should be constructed with this in mind by, perhaps, using ballast weight appropriately. The use of virtual scopes, who don't weigh anything, may also cause Soldiers to notice – those, too, can also be accounted for using ballast weight attached to the weapon.

Pose

The noun used to describe position and the orientation of an object is *pose* ('6 DoF', a common term in this industry, can be considered its adjective equivalent). The pose of the HMD is used to align the virtual objects with the live objects, so that they appear integrated both visually and acoustically. Pose needs to be captured at a rate that is at least equal to or greater than the visual frame rate of the HMD for the best performance – otherwise the virtual objects may appear to drag with the live objects in the scene. There are two approaches in calculating pose are:

- Outside-In. This utilizes an off-Soldier infrastructure (e.g., cameras on support trusses) to track the pose of the Soldier's HMD. This method minimizes what the Soldier needs to wear, which could be nothing more than a reflector this size of a dime. However, this requires an extensive infrastructure to cover the entire training area – and must account for all angles to ensure that there is no occlusion between the infrastructure cameras and the HMD. This infrastructure must be carefully installed and calibrated insomuch that typically only rectangular, wide-open areas are only feasible for use – which is a far cry from a MOUT. Also, the infrastructure typically relies on using infrared light, which is rendered helpless in sunlight.
- Inside-Out. On-Soldier technology uses environmental data to determine the pose of the Soldier's HMD. A common method uses a camera(s) embedded in the HMD that uses one or more simultaneous location and mapping (SLAM) algorithms. SLAM relies on finding an array of static points in a live scene to then calculate pose of the HMD – even with a single camera (Davison, 2002). This method *nearly* never requires any alterations to the training infrastructure (e.g., MOUT). Any alterations would involve intentionally placing inconspicuous objects in the environment for the SLAM to operate, but only if the environment was devoid of enough points, which is very rare. Note that this is not the same possible deficiency that can be found with depth-sensing methods like stereovision – far fewer points are typically needed with SLAM. In our experience, SLAM can be lost if the environment is nearly pitch black in darkness because it can no longer find enough points. Also, if there are bright flashes of light, the camera could be blinded because of its insufficient dynamic range from its camera(s). Finally, the SLAM relies on levering static points: If too many points in the SLAM camera(s)'s FOV are moving, its performance can be degraded.

We prefer inside-out pose calculation because, when compared to outside-in solutions, it greatly simplifies the setup, reduces costs, and is more reliable with tracking in any environment. The deficiencies of SLAM are rarely observed in dismounted Soldier training. There are also a multitude of pose recovery methods that are used if the SLAM is ever disheveled. Last, having control of the exercise means control over the issues that may arise. For example, the issues with nighttime can be minimized by using pass-through video HMDs: These can make daytime appear as nighttime because they do not have an LTL-FOV and they can render opaque objects – allowing for, say, completely covering a sunny sky with a virtual starlit night. Thus, the exercise can be run during the daytime but Soldiers will only see nighttime.

Seven of the eleven HMDs we evaluated had integrated outside-in SLAM-based pose tracking. Two others were currently working this into their near-future products.

Tracking Body Parts

Tracking body parts serves a minor role in AR versus VR. In VR, we track body parts so Soldiers can see each other's body movements, such are arm/hand signals, crouching, or going prone. In AR, Soldiers can see these actions in the live environment by default. This reduces the need to track body parts for purposes of:

- Cover (e.g., to calculate P_k ratio)
- Concealment. For example, virtual OPFOR reacts to "seeing" a hidden Soldier's arm move. The motion of the arm must be tracked so the underlying game engine – which controls the actions of the virtual OPFOR – can react to it.
- Providing the ability for BLUFOR semi-automated forces (SAF) to react to arm/hand signals. This kind of functionality is then largely dependent upon making such BLUFOR SAF a reality in an AR environment, which has yet to materialize at the time of this writing.

What does body part tracking have to do with HMDs? Several HMDs are steadily moving down the path of using their own on-board depth-sensing to detect the user's body movements. This is already seen to some degree with gesture detection and could possibly be increased to other limbs with wider FOVs of the HMDs' tracking methods (which vary widely). Widening the FOV may not result in complete limb-tracking. A research effort that resulted in the EgoCap shows exceptional limb-tracking, even with occluding conditions (Rhodin et al, 2016). Prototypes of this device are shown in the picture to the right – a form factor that would not be suitable for dismounted Soldier training because of the likelihood of collision with the Soldiers' surroundings, among other issues.

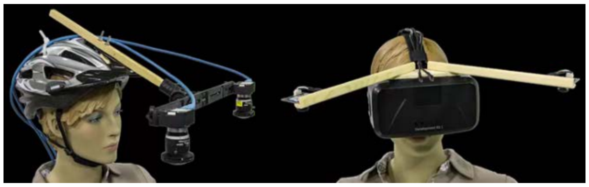
Figure 5. EgoCap prototypes. Courtesy: EgoCap.

Six of the eleven HMDs we evaluated supported some kind of gesturing. None of them, however, could track anything more than the lower arms and hands – and it was only when these body parts were within the FOV of the HMD.

Mixing of Live and Virtual Sound

Because many HMDs are integrating audio into their units, we felt it prudent to address some considerations with mixing live and virtual audio together. Proper audio cues is part of the fidelity required in AR-based training (Livingston, Brown, Swan, Goldiez, Baillot, & Schmidt, 2005). We begin with HMDs that provide sound externally (i.e., no headphones). The use of external audio is problematic because it will cause other Soldiers in close proximity to hear the same sounds in multiples – like an echoing effect. Furthermore, we weren't able to find any external audio HMDs whose bass was deep enough for battlefield effects or the guttural noises produced by nearby (virtual) passing armored vehicles.

This leads to the use of headphones, however, the headphones must be utilized so that live audio can naturally mix with virtual audio – which is counter to the consumer design of headphones, whose goal is to keep live noise out of the user's ears. Strangely, our qualitative testing revealed that very inexpensive earbuds (e.g., such as those bought on airlines) allow a pleasing amount of live audio to enter the ear. We realize that many HMDs are tethered (e.g., to a backpack computer) and headphones can also be tethered as well; however, minimizing the use of wiring across the Soldier's body reduces the chances of snagging, encumbrance in donning/doffing the equipment, and reduces aesthetic qualities. Last, we acknowledge that everyone will perceive the same audio slightly differently because of differences in ear/head structure through the head related transfer function or HRTF (Duraiswami, n.d.). However, crafting audio specifically to each Soldier's biology does not translate to better performance over a generalized form of HRTF (Muller et al., 2006), so headphones or earbuds that don't account for precise HRTF should provide realistic enough audio for the Soldiers.

All tethered HMDs supported the use of headphones by virtue of their accompanying tethered computer. All tetherless HMDs we evaluated also support headphones had a 3.5mm jack and/or a USB port. There are adapters that can provide between 3.5mm-to-USB but it's recommended that these should be avoided because this adds form factor – albeit small – that will add some degree of physical encumbrance to the Soldier.

Shading and Shadowing

Part of placing virtual objects realistically into AR is having the ability to naturally provide shading and shadowing on these objects (Livingston et al., 2005). In cinema, the ability to integrate virtual objects into a live scene is very difficult in a large part because of shading and shadowing. Movies are able to accomplish this through extensive post-processing techniques that aren't currently available for the real-time demands of AR. Optical see-through HMDs have an immediate disadvantage because they cannot show black pixels because all pixels rendered are additive – they can only add light to the display – objects will look "ghostly" (Livingston et al., 2005). Diming the ambient light does not solve this problem as virtual objects either disappear because of their transparency or stand out too in an attempt to overcome this transparency problem. Pass-through video HMDs, on the other hand, can show all ranges of shading and shadowing because they can show black pixels. Pass-through video HMDs or any HMD that can show black pixels are a necessity to naturally show shading and shadowing of virtual objects.

None of the seven optical see-through HMDs are recommended. This was true with HMDs with tinting or photochromic lensing, which provide an all-or-nothing approach to how the entire scene should be shaded. To be fair, shading was not the vendor(s)'s point in using tinting or photochromic lensing – the point was to allow for the HMDs to be used outside in ambient sunlight or in brightly lit indoor facilities. All four of the video pass-through HMDs can provide this capability, if the system can provide it with the appropriate information to cast shading and shadowing properly on both live and virtual objects.

Physical Encumbrance

Physical encumbrance is any burden the HMD will place on the Soldier in performing physical tasks. The Soldier will be wearing more equipment than they take into combat simply by the virtue that they're using an HMD, which is not a tactical device. So, which HMDs are the most capable of minimizing physical encumbrance? The HMD is either tethered to an external computing device(s) and/or it is tetherless – either of these methods may result in a form factor that occupies more space than any current Authorized Protective Eyewear List (APEL) eyewear ("Authorized Protective Eyewear List," 2017). By example, consider a fireman's carry: A technique to carry an unconscious or severely comrade over a Soldier's shoulder (see ATP 4-25.13). An HMD that is larger than APEL-approved eyewear may come in contact with the carried Soldier – and this would impede training.

Tethered HMDs have pros and cons. The pros are that they will typically have more computing power and this computer power is upgradeable. The cons are that the wire(s) that tethers the HMD to the computing device can get tangled or snagged while in use. Also, if the wire(s) are too taut, it will impede free movement of the Soldier's head. If the tethered computing device is a backpack computer, the Soldier won't be able to wear his/her own gear, such as a MOLLE backpack. If the tethered computing device is much smaller (e.g., something that would fit in a butt pack), it would mean that the Soldier could don more of his/her own gear, however, the processing power will be far lower than the backpack computer.

Consider HMDs with the ability to track body parts as a way of avoiding have to use on-limb motion trackers. On-limb motion trackers add to encumbrance and the overall complexity of the solution. From our experience with DSTS, Soldiers put up with them out of necessity, but would have rather not needed them in the first place.

It is worth mentioning here the benefit of light field HMDs as a way of reducing encumbrance. These allow the Soldier to experience far less discomfort and visual fatigue when trying to focus on objects at varying depths. This is because of the vergence-accommodation conflict that exists with non-light field HMDs (Kramida & Varshney, 2016; Lanman et al, 2013).

RECOMMENDATIONS AND CONCLUSION

It is likely that any HMD selected will have noticeable deficiencies beyond what a vendor will publicly state or even be aware of. One should take great care in examining beyond the commonly stated characteristics such as FOV or weight. We recommend that as many HMDs as possible should be evaluated hands-on against specific, common challenges that the user will encounter. The figure below demonstrates the difficulty in making informed decisions from only using public information (e.g., protrusion distance is not disclosed at all). Walk through the tasks that the Soldier (or any user) will have action-by-action, at the most atomic level possible to better discern which capabilities are must-have’s and which ones are not. Last, challenge the vendors on their future plans to ensure that deficiencies have a path to be rectified – many of them are glad to hear what your needs are.

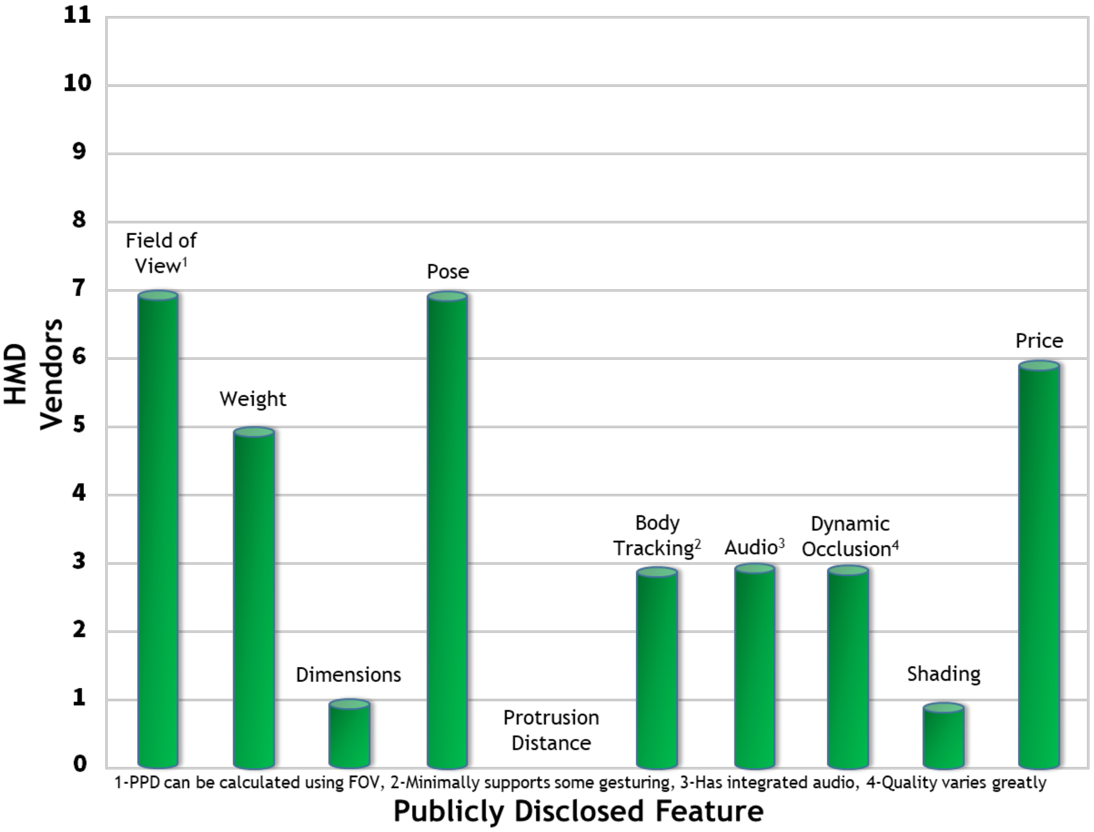

Figure 6. Publicly disclosed HMD features

REFERENCES

Authorized protective eyewear list (APEL). (2017). Program Executive Office Soldier. Retrieved from: http://www.peosoldier.army.mil/equipment/eyewear/

Bailey, R., Wilz, S., & Arthur, J. (2012). Conceptual design standards for eXternal visibility system (XVS) sensor and display resolution. NASA. NASA/TM–2012-217340. Retrieved from: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120002665.pdf

Binstock, A. (2015, May 15). Powering the rift [Blog post]. Retrieved from https://www3.oculus.com/en-us/blog/powering-the-rift/

Carmack, J. (2104, Oct 12). Transcript from John Carmack's keynote at Oculus Connect 2014. Retrieved from: https://singjupost.com/john-carmacks-keynote-oculus-connect-2014-transcript/

Davison, A. (2002). SLAM with a single camera. Retrieved from University of Oxford, UK website: https://www.doc.ic.ac.uk/~ajd/Publications/davison_cml2002.pdf

Duraiswami, R. Introduction to HRTFs. (n.d.). Retrieved from the University of Maryland website: http://www.umiacs.umd.edu/~ramani/cmsc828d_audio/HRTF_INTRO.pdf

Endsley, M. (1995). Toward a theory of situational awareness in dynamic systems. Human Factors. 37(1), 32-64.

Enhanced night vision goggle (ENVG) AN/PSQ-20. (2015). Project Manager Soldier Sensors and Lasers. Retrieved from http://www.peosoldier.army.mil/portfolio/#117

FM 3-22.9 Rifle marksmanship M16-/M4-series weapons. (2008). Headquarters Department of the Army. Retrieved from: http://usacac.army.mil/sites/default/files/misc/doctrine/CDG/cdg_resources/manuals/fm/fm3_22x9.pdf

Freriks, I. (2012). Test method: How we measure dynamic range. Camera Stuff Review. Retrieved from: https://www.camerastuffreview.com/test-methods/test-method-how-we-measure-dynamic-range

Kramida, G., & Varshney, A. (2016). Resolving the vergence-accommodation conflict in head-mounted displays. IEEE Transactions on Visualization and Computer Graphics, 22(7), 1912-1931. doi:10.1109/tvcg.2015.2473855

Muscatatuck Urban Training Center. (n.d.). A State of Defense. Retrieved from: http://www.astateofdefense.com/muscatatuck-urban-training-center.html

Lanman, D., & Luebke, D. (2013). Near-eye light field displays. ACM SIGGRAPH 2013 Emerging Technologies on - SIGGRAPH 13. doi:10.1145/2503368.2503379

Lavalle, S. M., Yershova, A., Katsev, M., & Antonov, M. (2014). Head tracking for the Oculus Rift. 2014 IEEE International Conference on Robotics and Automation (ICRA). doi:10.1109/icra.2014.6906608

Livingston, M. A., Brown, D., Swan, J. E., Goldiez, B., Baillot, Y., & Schmidt, G. S. Applying a testing methodology to augmented reality interfaces to simulation systems. Proceedings from the 2005 International Conference on Human-Computer Interface Advances for Modeling and Simulation. January 23-25, 2005. New Orleans, LA: SIMCHI.

Muller, P., Cohn, J., Schmorrow, D., Stripling, R., Stanney, K., Milham, L., ... Whitton, M. C. (2006). The fidelity matrix: mapping system fidelity to training outcome. The Interservice/Industry Training, Simulation & Education *Conference (I/ITSEC).* Orlando, FL: National Training Systems Association.

National Institute for Standards and Technology. (2006). Standards for visual acuity. Newtown, CT: John M Evans LLC

Perfect pixel count. (2014). Axis Communications. Retrieved from: https://www.axis.com/files/feature_articles/ar_perfect_pixel_count_55971_en_1402_lo.pdf

Ren, D., Goldschwendt, T., Chang, Y., & Hollerer, T. (2016). Evaluating wide-field-of-view augmented reality with mixed reality simulation. 2016 IEEE Virtual Reality (VR). doi:10.1109/vr.2016.7504692

Rhodin, H., Richardt, C., Casas, D., Insafutdinov, E., Shafiei, M., Seidel, H., . . . Theobalt, C. (2016). EgoCap. ACM Transactions on Graphics, 35(6), 1-11. doi:10.1145/2980179.2980235

Stevens, J. & Kincaid, P. (2014). Measuring visual displays' effect on novice performance in door gunnery. The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC). Orlando, FL: National Training Systems Association.

troy_s. How to measure the dynamic range of an HDRi? [answered Nov 3 2016 at 19:43]. Message posted to: https://blender.stackexchange.com/questions/66503/how-to-measure-the-dynamic-range-of-an-hdri

Two new soldier optics work together to offer rapid target acquisition. (2015). *Soldier Systems*. Retrieved from: http://soldiersystems.net/2015/07/23/two-complimentary-new-soldier-optics-work-together-to-offer-rapid-target-acquisition/

Vandenberghe, JD. (2010). The eye. Provideo Coalition. Retrieved from: https://www.provideocoalition.com/the_eye/

Wen, Y. F., Wong, H. M., Lin, R., Yin, G., & Mcgrath, C. (2015). Inter-ethnic/Racial facial variations: A systematic review and Bayesian meta-analysis of photogrammetric studies. Plos One, 10(8). doi:10.1371/journal.pone.0134525

Zamora, P. (2013). Virtual training puts the 'real' in realistic environment. Retrieved from https://www.army.mil/article/97582/Virtual_training_puts_the__real__in_realistic_environment

Zankl, A., Eberle, L., Molinari, L., & Schinzel, A. (2002). Growth charts for nose length, nasal protrusion, and philtrum length from birth to 97 years. American Journal of Medical Genetics, 111(4), 388-391. doi:10.1002/ajmg.10472