AREA Augmented Reality for Enterprise Alliance

AREA Technical Report: User Experience Design for Enterprise Augmented Reality

Employing Augmented Reality (AR) devices to present context-driven information deserves very careful consideration. This is especially true in working environments, where presenting virtual elements (often referred to as “augmentations”) on top of real world objects might negatively affect the completion of the task at hand – or even compromise the physical security of the users.

This report provides a comprehensive view of the design process and a set of practical guidelines for choosing the appropriate AR presentation device and developing meaningful and intuitive user experiences (UX).

The Beginning, the Middle and the End

All too often, AR project decision-makers begin their projects with the selection of novel devices for AR presentation and then look for use cases. Fitting the user experience and use case to the strengths and limitations of a device is frequently the source of many challenges. In some cases, it is even at the root of AR introduction project failures, loss of confidence in the technology, and further descent into the fabled “trough of disillusionment” that is described in the Gartner Hype Cycle.

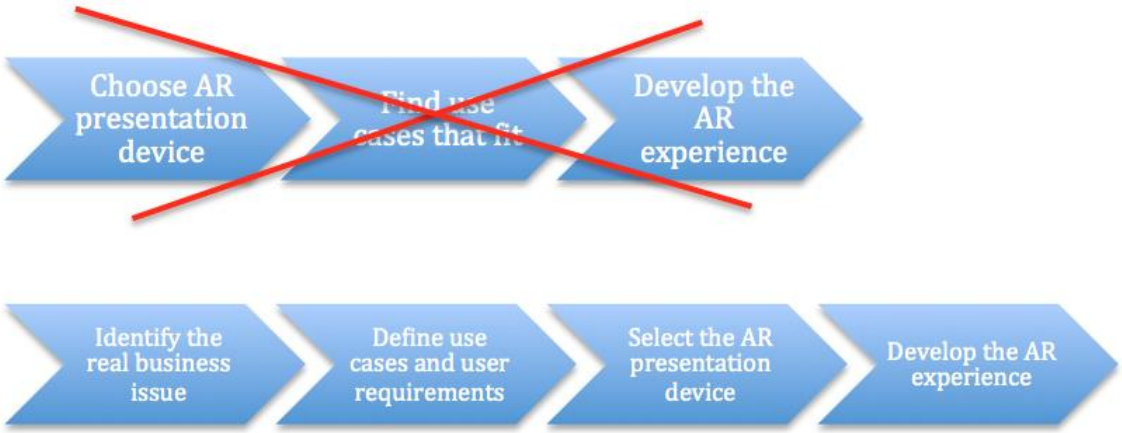

Figure 1: Increase likelihood of success by selecting the AR presentation device based on user requirements.

For successful enterprise AR projects - whether during introduction or with experienced groups - the process should begin with identification of a clear business need. Use cases are defined within the context of the needs. The team then defines the user’s requirements.

Once the use cases are identified and understood, often using a storyboard, the developer works with team members to determine the best AR presentation system components (e.g., hardware and software). Finally, once the AR presentation system is defined, the developer works within the technology constraints of the chosen system to design the optimal user experience.

Use Cases and Recommendations

We are surrounded by objects in the physical world, but most people don't need AR to interact with them. Enterprise Augmented Reality use cases must focus on those situations in which existing approaches for using or maintaining and repairing physical objects are known to have weaknesses; for example, situations where the data available is incomplete, or hard to follow.

With limited time and resources, early AR projects should also serve to illustrate the future – providing a glimpse of how many complex processes involving unfamiliar objects can be connected with digital instructions and other enterprise assets.

In this section, we discuss use cases that are frequently chosen for AR introduction in industrial settings. Use case analysis and description are an important step that helps designers to make decisions during the design process. There are several methods in the scientific literature that analyze the task for which the solution is to be designed and extract design requirements [1, 2]. Most of these techniques aim at helping the designer understand the constraints that need to be respected and the features that the user needs in the final experience, and the extent to which the features are a requirement. The essential components in these analytical techniques include:

1. Information requirements at each step
2. Environmental conditions during the task
3. Human factors and ergonomics
4. Desired outcome at each step
5. External constraints affecting the workflow

This report uses the five components above to describe a few sample use cases that are popular for AR in industrial settings. This will clarify the approach as it begins with a general analysis of the target task and then is further broken down along multiple dimensions.

Three general use cases will illustrate the recommended approach:

- Warehouse picking
- Assembling a new product
- Performing maintenance and repair operations

These three general use cases have been chosen because they are common AR introduction projects. The reader can easily tailor them further for unique settings and requirements.

The use cases then drive the selection of solution components across four categories:

- Presentation system
- Display technology
- Mobility
- Connectivity

Warehouse Picking

The scenario characterizing this family of tasks involves moving goods in a well-known facility. The typical blueprint for warehouse picking tasks includes:

- Navigation inside the facility towards the location of interest
- Identification of the item of interest
- Loading/unloading of the item
- Documentation of the action performed

Information required: In this scenario, the worker needs to be able to identify the place where the item is stored and how to reach it, the identifier of the item (usually an alphanumeric code or a barcode), and the action to perform, including the relative destination, if appropriate. Knowing the progress status in the list of actions to perform can also help the worker with timing and process understanding.

Environmental conditions: It is safe to assume that these tasks are performed in facilities with reasonably good lighting conditions, dry atmosphere, and normal room temperatures. However, the level of ambient noise can, in some cases, be relatively high due to forklift and other machines operating around the user.

Ergonomics: Workers need the use of both hands most of the time. Despite the fact that items of interest are usually large and easy to scan with any device, the code or identifier can be in places where access is difficult or impossible.

Desired outcome: The worker needs to pick a list of items and perform actions to completion on the relevant order, then to document the result using the appropriate forms.

External constraints: There are few if any external constraints affecting warehouse picking use cases.

Assembling a New Product

During assembly scenarios, workers use mechanical and electric components following a standard procedure that guides them through the assembly of an object composed of multiple parts. These components might have different sizes and can be functional parts or only serve to hold critical parts together as part of the final object. The procedure can require using specific tools.

Information required: At each step of the task, the worker must focus on the part that needs to be assembled, its exact location relative to where and how it should be in its final position, the space in which the worker will perform the assembly movements, and the tools necessary to carry out the action.

Environmental conditions: Assembly tasks are usually performed indoors. The entire task is normally located in a specific place that is dedicated to the assembly of that component or of a known category of components. Lighting conditions are controlled while noise level can be high depending on the machines operating in proximity.

Ergonomics: The tasks require the use of both hands at all times. The amount of physical movement required varies but is frequently limited as the actions are performed around a stationary object.

Desired outcome: At each step, the worker verifies that the components are assembled in the correct configuration.

External constraints: In some cases, the task may have time constraints related to the actions performed (e.g.., a part needs to cool down after welding before being able to operate).

Maintenance and Repair Operations

Maintenance and repair operations (MRO) are procedural tasks that involve the revision of the status of a piece of equipment, the diagnosis of a problem and, when necessary, the repair of an identified fault. MRO procedures can be performed on any type of component and in potentially any location. The procedure often follows these steps:

1. Receive notification of a maintenance order
2. Identify the location
3. Identify the object
4. Diagnose the status of the object
5. Identify the fault
6. Perform the repair procedure
7. Report that the procedure has been performed successfully

The repair procedure can include the replacement of a part of the equipment and the disassembly/reassembly of many adjacent parts.

Information required: The user needs to be guided to the location of the object in need of maintenance. Once at the location, the user needs to identify the equipment. The user must retrieve all the information needed to diagnose the problem with the machine and be aware of the procedure to perform in order to perform the repair.

Environmental conditions: MRO procedures can be carried out in any conditions. Lighting, temperature and humidity conditions can vary, both in indoor and outdoor settings.

Ergonomics: This category of operations is characterized by the high mobility needed by the workers. Procedures often require the use of both hands and are sometimes performed in spaces that do not permit free body movement.

Desired outcome: The task must be completed following the procedure for quality assurance to ensure that faults are identified and fixed. Lastly, the user must report the completion of the task to the central system.

External conditions: The successful completion of MRO procedures can depend on the availability of diagnostic sensor data related to the equipment under inspection.

Hardware Considerations

The second step to bear in mind during the design phase of a project involves identifying technological requirements related to the task. Frequently, a design might look promising and lab trials might support the choice of hardware, but when the technology is deployed in the field, technological limitations arise that hinder the effectiveness of the tool, thus forcing the design process to restart. These delays and costs associated with incorrect choices can be avoided with a careful analysis of the technologies at the user's disposal and the constraints that the tasks and the environment pose on these technologies. The next sections provide examples of some of the important aspects that an AR designer needs to consider when choosing the most effective AR presentation hardware.

Presentation Systems

One of the most important decisions is the choice of device for presenting the AR experience. Currently, the options include:

- Handheld devices: The rear camera of tablets and smartphones captures the environment in front of the user while the display facing the user presents the AR experience (video see through). The touch screen is the primary interface for interaction with the user interface and the content.
- Wearable devices / Head-mounted displays (HMD): Camera-equipped smart glasses and visors display information directly in front of the eyes of the user without the need for him to hold a device. Industrial-grade systems in this category are often designed to be attached to or to replace hard hats or industry-complaint safety goggles. In some cases, these devices are tethered to an external controller that serves as the interface and provides supplementary battery power.
- Stationary: these devices are characterized by a larger form factor and are statically installed in a fixed physical location. Stationary AR systems are composed of a camera system capturing the environment around the object of interest and a display device. Display devices normally used in stationary systems are based on projection technology or large screens mounted in easily accessible locations. In the first case,

projectors are installed above the object of interest and project light in shapes, text and images directly on the object or the surrounding surfaces. Alternatively, the screen displays the live video feed captured by the camera, blending in the overlaid information, similar to a handheld device but mounted on a bracket or arm.

When choosing among these presentation hardware options, some considerations are crucial to take into account at the start of the design phase.

First, these devices clearly offer different capabilities in terms of mobility. While handheld devices and smart glasses are highly mobile and can be easily carried around, stationary systems are strictly tied to a precise location, making them unsuitable for tasks that require roaming or performing tasks anywhere other than the station for which they are designated. Consequently, if a task is somewhat mobile (i.e., roaming inside a facility) and highly mobile (i.e., tasks in remote locations or outdoors), either of the first two options can be considered. On the other hand, stationary systems are particularly effective in situations where the task is always performed in the same place (i.e., the assembly of a mechanical part) as they do not force the user to wear or physically interact with an external device.

Figure 2: Tablet AR (on the left) and projection-based AR (on the right - Image credits Marner et al. [3]).

Another factor that influences the choice of AR presentation hardware is the user's need for hands to perform a task. Both smart glasses and projection AR systems are hands-free AR presentation systems, while smartphones and tablets need to be held as they are pointed at the target object in order to allow the camera to capture the scene and overlay information. This may be inappropriate for tasks that require simultaneous use of both hands when using an AR experience.

Display Technology

Display technologies also play an important role when organizations are choosing hardware. The display technology options include:

- Video See Through (VST): a display screen shows the video feed being captured by the camera while the augmentations are rendered on top of it, making it appear as if the virtual objects were embedded into the physical environment. The user perceives their environment while looking at the display. VSTs are used for both handheld devices and smart glasses.

Figure 3: Example of video see-through AR on a smartphone. Source: Wikipedia

- Optical See-Through (OST): the display is composed of a semi-transparent support on which the augmentations are rendered and reflected into the human eye. The position of the augmentations in space is calculated based on the video feed captured by the device camera

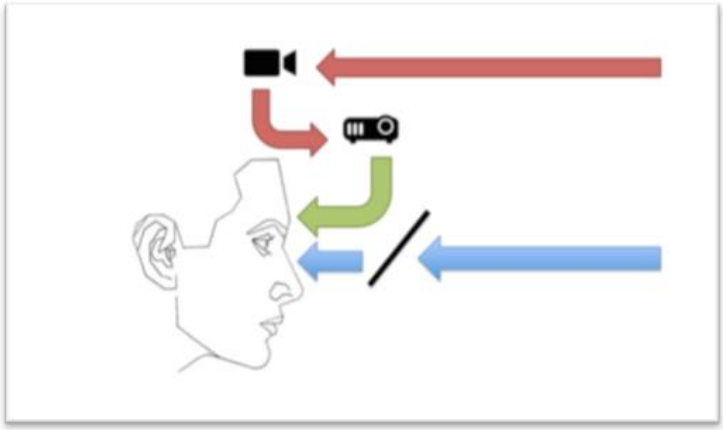

Figure 4: Conceptual architecture of OST displays.

The user perceives the environment through the semi-transparent glass while also perceiving the augmentations. These displays are used for smart glasses.

Figure 5: Example of an optical see-through display for an aircraft windshield. Source WIkipedia

- Projection: A projector renders the augmentations directly on the target objects in the form of monochromatic or color shapes. The user perceives the environment and the augmentations directly without any mediation.

Figure 6: Example of projected AR on a real car. Image credits [4]

These three technologies provide very different experiences as they are constrained by their individual technical limitations.

Video See-Through display technologies are the most diffuse as they are available on most commercial grade devices, such as tablets and smartphones or camera-equipped virtual reality visors. The use of this technology for AR experiences introduces a delay between the user's movements in space and the final rendering of the same scene with the blended overlays. This is due to the computational process involved in the creation of the AR scene. Once a frame is captured by the camera, it is processed for recognition and tracking. Once the camera pose spatial transformation is calculated, the AR engine calculates the relative spatial transformation of the overlaid augmentations. Finally, the rendering engine blends the augmentations in the frame captured and renders the result on the display. Current technologies allow for a delay of as little as 200 milliseconds, which is, however, large enough to be perceived. In fact, the latency between the user's movement and the delayed movement of the scene observed on the display has been demonstrated to hinder the experience, causing confusion [5] and, when used for wearable displays, a sense of nausea called "cybersickness" [6].

Guideline: for video see-through display-equipped smart glasses, experiences do not have to force the user to wear the device for long sessions.

In addition, the form factor of headsets requires the FOV to be completely immersed in the scene displayed, in order to avoid the displacement caused by the different scenes observed with peripheral vision. This might be counterproductive and dangerous in situations where the user needs to be constantly aware of the surroundings.

Guideline: avoid video see-through smart glasses for tasks that require the user to have full spatial awareness of the surrounding environment.

Even with these limitations, VST is currently used to deliver most of the handheld AR experiences: the hardware needed is already embedded into tablets and smartphones and the user is already used to holding camera-equipped devices and pointing them towards an object of interest (the same interaction can be experienced when taking a photo), observing the environment through the “digital window.”

Guideline: camera and display resolution should be as close as possible.

In many cases the resolution of video feed is lower than the display resolution as frame rate is preferred to quality during video capture. This may produce a strongly visible mismatch between a high-resolution overlay and a low-resolution background. A solution usually lowers the rendering quality.

Optical see-through technologies are, in many cases, not affected by the peripheral FOV obstruction, as the user can perceive the environment through a transparent material. These displays are usually employed on smart glasses because they are completely unobtrusive, allowing the wearer to directly perceive the environment. For this reason, this is the most popular solution adopted for wearable devices. The possibility of overlaying virtual imagery on the environment around the user makes this category of displays the most sought after among both providers and users of AR. However, significant technological impediments limit the applications of these displays.

The first important limitation affecting optical see-through technology is the limited portion of FOV that can be used to render virtual imagery. The current state of the art does not allow the manufacturing of lenses capable of displaying virtual imagery in a FOV comparable to what the human eye usually perceives (more than 180 degrees).

This creates some difficulties in the application design process. During tasks that require the visualization of a large amount of information, the user is forced to “scan” the environment during the visual search of the needed information in order to compensate for the small FOV. The head movements required to visually scan the environment are not only ergonomically inefficient – visual search through eye movements is ten times faster – and uncomfortable, but also are known to decrease the human ability to acquire information during problem-solving tasks [7]. Nevertheless, UX experts are trying to implement design solutions that aim to compensate for this limitation.

Guideline: divide information and overlays in small self-contained chunks if the content will be delivered on an OST display.

Guideline: display resolution affects the size of the augmentations. Objects too small will not be visible/readable with a low-resolution display.

Designing the AR content and interface as separate visual objects helps the user in identifying entities. If the overlay exceeds the useful FOV of the display, the user will be forced to step back to visualize it or turn, potentially missing parts of the information if the content is dynamic.

Guideline: the AR experience should provide visual cues to direct the user’s attention if the FOV is limited.

Figure 7: A red visual arrow can direct the user's gaze direction.

The augmented scene is observed through the small window created by the OST display. Due to the limited FOV of this window, in those scenarios where the user has to

look towards a relevant spatial element (virtual or otherwise), the interface can guide the gaze of the user towards that content using spatial cues.

Background and light conditions heavily affect the visibility and readability of AR content when visualized with OST displays. Current OST displays are able to render content with a maximum of 80% opacity. This means that augmentations blend with 20% of the light reflected by the environment. Consequently, bright environments affect the visibility of the virtual imagery. The same concept applies when considering the background color. Low contrast between the background and the augmentations hinders their visibility.

Guideline: avoid optical see-through displays for very bright environments and outdoors. In addition, test the designed experience in the environment it will be used in order to check color consistency.

Another important factor to consider is the difference between monocular and binocular displays for smart glasses.

On the one hand, monocular HMDs occupy a portion of the FOV of only one of the two eyes. Also called “assistive displays,” these devices are implemented with both VST and OST technologies. On the other hand, binocular displays occupy the same portion of FOV for both eyes. The main difference between these two categories is that a binocular display is able to deliver stereoscopic augmentations. Therefore binocular displays are the only ones able to deliver spatially registered AR (also known as Mixed Reality or MR). Nevertheless, monocular displays are far less obtrusive then binocular ones and, being placed in the peripheral vision of one of the eyes, can be easily scanned with a simple eye movement.

Guideline: use binocular displays if spatially registered AR is part of the UX. If only checklists or contextual information needs to be displayed, the experience might benefit from a monocular display.

Projection technologies have a very different set of characteristics. Firstly, by nature, projected images cannot render stereoscopic content. Therefore, it is not recommended to make use of projectors for displaying overlaid 3D content. This heavily limits the number of applications appropriate for projection technologies. Projection AR (also called Spatial AR) is commonly used to highlight points or objects of interest, or to render small amounts of text around them and, in some cases, geometrical shapes or schemes.

Lighting conditions also affect the readability of projected images. Although modern projectors render images that are also visible in strong environmental light, sudden variations in light conditions can hinder the readability of text (the human eye requires several milliseconds to adapt to a different luminance level).

Guideline: when using projection technology, control lighting conditions though luminosity sensors and lights.

When designing content and interfaces that are going to be projected on objects, it is important to generate a good representation of the surfaces on which the shapes will be projected. In fact, because these shapes are made of light beams, if the surface is irregular, protruded or reflective, the light will not be normally reflected and the shapes will appear distorted.

Guideline: make sure that projection surfaces are regular and non-reflective.

Connectivity

Most AR applications rely on the presence of some form of connectivity on the device they are deployed on, in order to download content, perform indoor positioning and upload real-time results of operations.

Technologies commonly used are:

- Wi-Fi IEEE 802.11 a/b/g/n
- Bluetooth 2.0 / 3.0 / 4.0 / LE
- 3G, 4G cellular networks

These technology standards vary in relation to the data transmission speed, type of data allowed and range.

Bluetooth has a very short transmission range (up to 20 meters) and is generally used for indoor positioning (combined with beacons) and non-data-intensive device communication. Wi-Fi connections are usually characterized by high-speed data transmission rates and a maximum transmission range of 50 meters, allowing usage for indoor applications. The extended range of 3G and 4G networks justifies their use for AR applications outdoors and in remote facilities that cannot provide Wi-Fi connections. However, the data transmission speed of these connections can vary according to atmospheric interference and signal coverage.

Network Connection Specs		
Technology	**Download peed (Mbit/s)**	**Range**
Wi-Fi a/b/g/n	10 - 600	~30m
Bluetooth 2.0/3.0/4.0/LE	1 - 4	~10m
3G	20 - 100	~1500m
4G	40 - 150	~1500m

Table 1: Options for network connectivity for AR experience delivery.

Network availability and data transmission speed can heavily affect the experience and the efficacy of AR applications: downloading the relevant technical documentation in real time is essential in many AR applications. For this reason, it is important to investigate in advance the availability and reliability of the connections that will be used, and design the content accordingly. For example, it is recommended to design low-resolution versions of images, textures and 3D models if these are used on slower networks for faster rendering.

Guideline: downloadable content size and resolution should be adapted according to expected network conditions.

Guideline: cellular networks can have highly variable transmission speeds. Data loss and delays can affect the validity of real-time data.

Industrial Standards

Organizations planning AR implementations must also consider prospective devices' characteristics and compliance with industrial safety standards. For example, worker health and safety codes stipulate that users are required to wear protective eyewear in some environments. In other industrial settings, the user may be required to wear a protective helmet, which could interfere with an AR device.

There are other industrial standards that apply to other settings, such as the need for sterilization in some food or pharmaceutical manufacturing plants. In oil and gas

environments, devices must be certified for safety (defined as meeting intrinsically safe requirements as in standard UL 913).

UX Design Considerations

Some considerations and recommendations in this section are driven by requirements introduced in the previous section. For example, the UI and interaction design are tightly constrained by the technological limitations of the hardware platform on which the UX will be delivered.

Information Display Space

Content for AR experiences is characterized by different requirements than content for traditional media, such as computers and tablets. The visual information is not constrained to the borders of the display frame, but is all around the user as part of the environment. Ideally, this liberates the information from the limitations imposed by the size of the screen, giving the designer more freedom to manipulate content towards the desired experience. For example, increasing the size of an overlay can highlight its importance in relation to the surrounding visual augmentations or real objects – a bigger arrow can signal task priority in dealing with an object. However, the practicalities of the current technologies limit what can be actually achieved. For example, the size of virtual objects needs to be enlarged if the display resolution hinders visibility or readability of information, but cannot be over-expanded due to the restricted FOV of many AR-enabled devices.

Guideline: content for use in AR experiences needs to be large in size (compared to what is designed for high-resolution tablets or other screens) and other steps may need to be taken (e.g., animating a simple asset) to ensure the information is clearly visible in the user's environment. In addition, crucial parts of the information should be simultaneously visible without users needing to turn their heads (in the same FOV).

Content positioning plays a very important role in AR experiences. The location of digital content can be conveniently mapped in two spatial reference systems [8]:

- Egocentric: also known as *user-centric space,* this reference system has its reference point in the user's body and, specifically, the center of the user's perceptual system, often identified with the head.
- Allocentric: also known as *object-centric space*, this reference system uses the object of interest as a reference point.

This distinction has been useful in scientific research to categorize the specific issues and solutions regarding each specific category of AR interfaces [9].

Egocentric interfaces are spatially organized around the user independently from the scene observed by the AR-enabled device, and can be accessed through the user's active query (e.g., turning the head right or left, looking up or down) – see Figure 8. In cases of handheld devices, these typically take the form of on-screen content that can occupy a portion (if not all) of the video feed, constantly overlaying it. HMDs can deliver egocentric content using two different techniques.

The most commonly used technique, known as Heads-Up Display (HUD), implies the placement of information in the peripheral portion of the useful FOV so that the content is constantly visible independently from the head movements. The second technique makes use of the inertial motion-sensing units (IMUs) installed in HMDs to place content on the sides of the user's head.

The user can visually query this information by rotating the head to one side or the other where the content is located. This technique assumes that there is a preferred head direction to keep clear from content, or the ability for the user to change this direction during usage (personal preferences). Egocentric interfaces are particularly useful when the user needs to access information while moving or changing environments, as it is not registered to any particular spatial reference point.

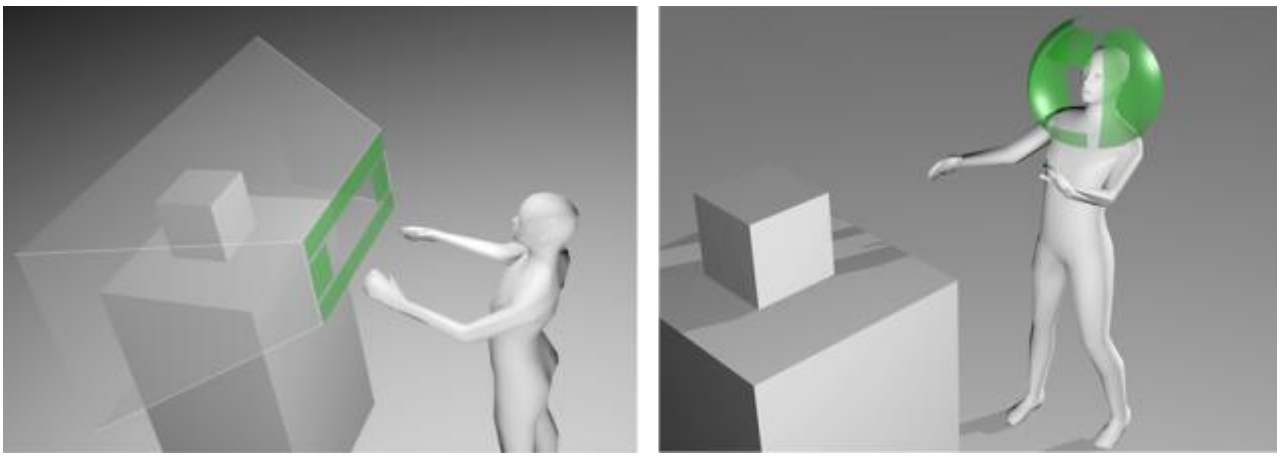

Figure 8: Reference space of HUD-like (on the left) and head-tracked (on the right) egocentric user interfaces.

There are also drawbacks to egocentric interfaces. For example, it is very easy to clutter the user's FOV with an overwhelming amount of information in the central part of the visual frustum, losing sight of important elements in the physical world and making the user uncomfortable.

Guideline: always-on user-centric interfaces should occupy a small portion of the FOV. Preferably, external corners and peripheral parts of the FOV are used to place content.

Although peripheral vision cannot be used to recognize patterns and understand the environment, it is particularly sensitive to variation in light modulation – i.e., a sudden change in luminosity [7]. This means that object movement (which changes the luminosity in the peripheral vision) is more likely to attract attention than shape or color change.

Guideline: object movement – bouncing, vibration, etc. – attracts the user's attention towards information notifications in the peripheral FOV better than other visual changes.

Object-centric content presentations are commonly used in AR user interfaces. Labels, text, 3D models, images and other media are overlaid on top of or around the object of interest. Their three-dimensional pose is independent from the viewer's perspective but estimated from the object's pose relative to the user.

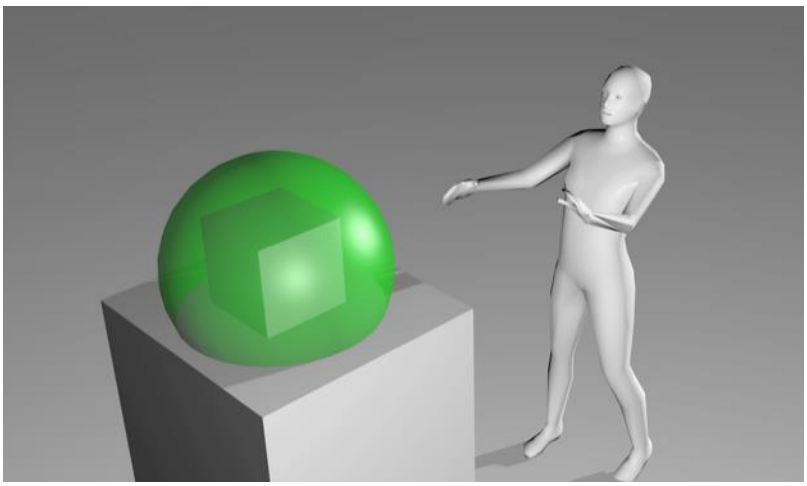

Figure 9: Reference for object-centric interfaces.

Object-centric interfaces are used to provide spatial cues about specific parts of the object or, more generally, to spatially identify the object of interest and to associate digital content with it. It sounds simple to provide object-centric content but in real-life scenarios, correctly designing the overlays that will be superimposed on the objects is not an easy task. Frequently, the experience is compromised by having too much or too little information of value attached to the object of interest. In many cases, AR applications automatically download information from enterprise databases. The amount of information is, therefore, not known *a priori* and the number of labels or text displayed can clutter the space surrounding the object, compromising the visibility of the surrounding space and the readability of the labels themselves.

Results of scientific research on these issues converge on several requirements when associating digital content with real-world objects [10]:

- Readability of digital content (labels and text)
- Non-ambiguous relationships between overlays and objects
- Minimal occlusion of other pertinent information and objects

Researchers have recommended the development of algorithms to manage the placement of text and labels in space and limit their number and size according to the amount of information available [11-13]. Today's commercially available solutions do not have such algorithms. The size and positioning of overlays in space is done manually. Future projects will need to provide automatic systems that first map the space surrounding the object and the user and then choose empty portions of the visible environment for placing digital content [11].

Guideline: the size and number of digital assets superimposed onto an object of interest must always be limited. If information about the environment is available during the design phase, position overlays appropriately and do not occlude surrounding relevant objects.

There are other known challenges. Object-centric annotations are, by definition, attached to a specific target. Their position with respect to the object may be inconvenient when performing tasks that require the user to change positions (high mobility). Some of these tasks require the user to perform operations on large machines or spatially distributed equipment. In these cases, users need to visualize equipment-related information independently of their physical locations.

Guideline: information can be displayed in user-centric perspectives if the task requires

high mobility and is spatially distributed. It is also useful for these tasks to allow the user to pin pieces of information from object- to user-centric display spaces, providing the ability to "carry" information in space.

Object-centric visualizations often are dependent not only on the visibility of the object itself, but also on the relative position and orientation of the user. Assumptions are made during the design phase about the user's perspective. The most common is that the user will be looking at an object from a specific position and angle – in front in most cases. If the digital content is designed to be static (spatially arranged in a fixed position), the experience designer needs to make sure that the user is correctly positioned in relation to the object when visualizing the overlays. This can be achieved by providing spatial cues to direct the user's point of view towards a position and orientation where the experience designer assumes the user to be [13].

Guideline: the user should look at the object of interest from the expected position and direction. If this cannot be guaranteed, spatial cues to guide the user towards a favorable position should be provided.

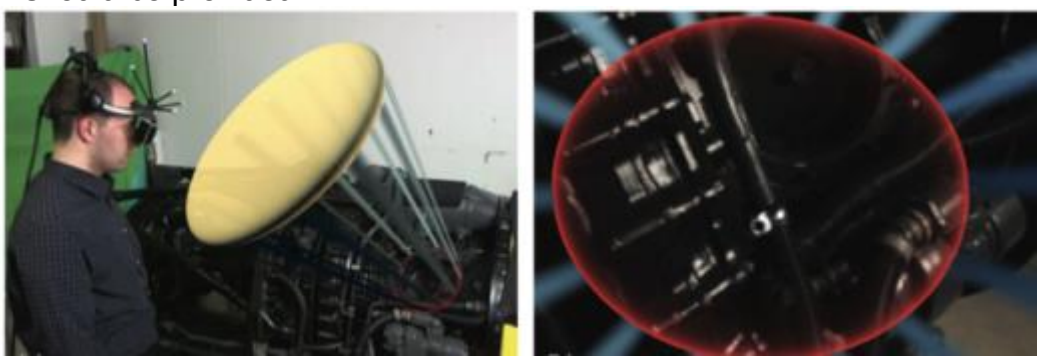

Figure 10: Parafrustum, a technique to direct the user towards the optimal point of observation. Image credits Sukan et al. [13]

Graphical Objects

The technological limitations and boundaries imposed by AR presentation systems impose a whole new set of requirements and guidelines for successful information and interface design.

In contrast with VR systems, where the experience is fully immersive, AR blends the real world and digital assets, suggesting the need for high levels of photographic realism. Such realism has been achieved with computer graphics special effects systems designed for television and cinema, but it requires post-processing of video streams in order to manually

or automatically overlay spatially registered digital information. Augmented Reality systems must process in real-time, therefore the degree of realism needs to be achieved in a different way or the requirement dropped or reduced.

Spatial annotations, for example, are a particular set of digital objects used to identify real objects in space, providing indications about the action to perform during the tasks and graphically labeling the environment. These annotations are superimposed directly on the object of interest and can often be combined with 3D virtual imagery and animations. Instances of this category of overlays are arrows, circles, color shapes or symbols. Spatial annotations can be designed to be rendered as 2D widgets positioned on a planar surface, or 3D images floating in the environment.

Figure 11: 2D content (on the left) and 3D content (on the right). Image credits [14]

Ideally, these two modalities of visual information presentation carry the same informative load and are equally effective for task instruction. However, scientific research demonstrates that the two strategies are not equally valid in every situation. In fact, while 2D shapes and symbols are clearer and more easily interpretable when rendered on objects with large planar surfaces, 3D arrows and shapes are more effective to annotate irregular objects and surfaces, as they more clearly provide depth cues and three-dimensional information [15]. In addition, 3D content is more computationally demanding than 2D shapes, and this affects system performance and battery life, both crucial requirements for industrial grade AR systems.

Guideline: flat (2D) content should be prioritized in order to improve performance and to prolong the battery life of the AR system. 3D content should be used only when necessary: 3D virtual objects are effective in providing depth cues.

Depth sensors allow AR systems to correctly estimate the spatial position of specific parts of the environment with respect to the target object on which the user will expect to see (and use) digital information. In addition, systems should be able to correctly position

the digital object with respect to the real world. Commercial graphics rendering engines can correctly display depth cues of two digital objects: when two digital objects are in the same position with respect to the user but at different distances, the object in the foreground will occlude the object in the background and objects at a distance will also appear smaller. This said, most commercial 2D tracking solutions in use on smartphones and tablets (such as Blippar, Vuforia or ARtoolkit SDKs) are not designed to correctly detect and correct for differences in real and digital object distances.

Since, in most cases, current AR presentation engines track the environment using only monocular RGB cameras, they do not have depth-mapping data. As a result, the rendering engines are unable to display only the visible portions of a digital object occluded by a real object based on any depth data. This may cause misinterpretation of spatial cues and mislead the user towards an object that was not intended for the experience. Without depth mapping technology integration, it is recommended to design experience content so that the spatial location of a digital object with respect to the target in the real world is unequivocally clear and the results do not take into account the user's depth perception.

Guideline: occlusion handling driven by depth sensing is a highly desirable feature in an AR presentation engine. However, if depth mapping is not available, the developer must design content so that spatial cues do not heavily rely on depth perception and occlusion awareness.

Successful color design also deeply affects the effectiveness of an AR experience. Especially for OST display technologies, the environmental lighting and background color heavily influence the color perceived by the user, as mentioned in the Display Technology section. It is therefore important to make sure that colors are perceived correctly by the user. This is done by tuning the color palette. This is especially important for those tasks during which the user needs to match virtual object to real ones – e.g., cable wiring by color.

Guideline: use high-contrast colors to enhance overlay visibility and text readability.

Guideline: take into account the background color and how it modifies the color perceived by the AR experience user.

Guideline: if the experience will be delivered on multiple AR presentation systems, consider the technology involved. While VST displays can render a wide range of colors, OST and projection technologies have limitations to consider – e.g., black is rendered as transparent in OST display, so it cannot be represented.

Interaction

User interaction with digital content in an AR experience is influenced by a number of factors: type of task, presentation device and environmental settings. The developer's choices will greatly affect system usability and, by extension, the project outcomes.

The common interaction techniques for AR systems are shown in Table 2.

Gesture

Gesture recognition can be obtained using a camera-enabled device that constantly monitors the scene searching for pre-defined hand movement patterns (i.e., gestures). Either RGB cameras combined with computer vision algorithms or depth-sensing cameras can be used to recognize gestures. Gesture-based interaction is based on the assumption that the user is aware of and uses only a predefined set of gestures, and how these gestures are interpreted by the AR system.

Guideline: provide help and training for gesture interaction in order to reduce delays due to any learning curve.

AR Interaction Techniques	
Interaction Technique	**Description**
Gestures	A camera system detects user gestures and translates them in machine input.
Speech	The user expresses a vocal command from a set of pre-defined phrases to interact with the system.
Touch	Touch-enabled devices like tablets and smartphones allow the user to touch the screen to interact with the system interface and the AR overlays.
Arm-/belt-worn devices	External devices equipped with buttons or touch sensors are programmed to enable user input.
Stare/gaze navigation	This technique tracks the user's head or eye movements to interpret the interaction intention and act accordingly.

Table 2: Common interaction techniques for AR systems.

It is important to limit the number of possible gestures and to use gestures that are simple and easy for the user to perform in the environment. If the interaction is difficult to perform and the gestures are complex or unnatural, the system could have difficulties recognizing them, leaving users frustrated and reducing the likelihood that the user will adopt the system. In addition, the developer should study the system's gesture recognition algorithms to determine if they are sufficiently robust to recognize the proposed gestures reliably in the proposed conditions.

External factors that can influence the effectiveness of recognition algorithms include:

1. Extreme lighting conditions
2. Reflective surfaces in the surroundings
3. Gloves worn by the user

Guideline: design small simple gestures for user interaction. Test the interaction design in the expected user environment to ensure that external conditions do not affect robustness of recognition.

There are other considerations for this interaction technique. Because the system has to continuously execute advanced computer vision algorithms to achieve a successful interaction, the computational load on the CPU or a dedicated processor is very high. This makes gesture recognition highly power consuming. Although this may seem an issue with which only hardware designers need to deal, it is crucial for UX designers to take the battery life of the device into consideration.

Speech

Speech recognition is another option for AR interaction. The microphone-equipped device constantly captures the audio surrounding the device while the audio stream is processed by an algorithm that converts the recognized speech into text. The text is then analyzed to identify words that match patterns from a pre-defined set of allowed commands. The advantage of this technique is that it is fairly easy to implement. Online services, such as the Google Speech to Text API, provide ready-made interfaces to enable it – and it is easy for most users to learn. In fact, many modern operating systems provide built-in voice command interfaces and users are already familiar with this type of interaction. In addition, while gesture interaction can be ineffective and frustrating in cases where the user needs to operate constantly using both hands, speech interaction enables a completely hands-free experience.

Ow them

Guideline: implement speech interaction for hands-free experiences.

There are some limitations to speech as a technique for system control. First, the robustness of speech recognition can be heavily affected by the noise level of the environment. Despite the fact that many modern microphones have noise canceling algorithms, the reliability of speech recognition is compromised by noise levels that overcome the user's voice. These unwanted sound sources are usually generated by heavy machinery present in the environment, but they can also be the result of the conversations of other workers (which can be interpreted as valid commands). In any case, by using a microphone array it is possible to locate the position of the sound source and disregard those that are not created by the authorized users.

Guideline: avoid speech recognition in noisy environments.

Another important consideration when implementing speech interaction is to make sure that the user is aware of the commands allowed and how to properly formulate them. This technique involves vocal or auditory interactions between the user and the device. The device does not display visual cues of the possible commands that the user can choose from – i.e., a button with the text "Next" suggests that tapping that button will most probably make the session move to the next phase.

Guideline: display hints and suggest commands that can be used, especially for novice users.

Talking is an activity performed daily and natural language has a complexity that goes far beyond what a simple command recognition algorithm can process. If complex natural language processing techniques are not implemented, the algorithm may fail to recognize the user's command even if it is properly formulated. For example, a user could say "select the second item" followed by "and the two items below." Understanding such commands implies a complex analysis of the contextual relationship among words that many speech processors do not support.

Guideline: instruct the user not to give complex instructions using natural language.

Touch and External Devices

Touch input is a familiar mechanism for users and developers. As most AR-enabled handheld devices offer multi-touch input by design, it is very easy to implement touch-based

user interfaces and to design touch-based content manipulation. However, handheld devices are not the only devices that can provide touch-based interaction. Research prototypes have demonstrated how it is possible to develop finger-based interaction with projected overlays on planar surfaces using computer vision [16].

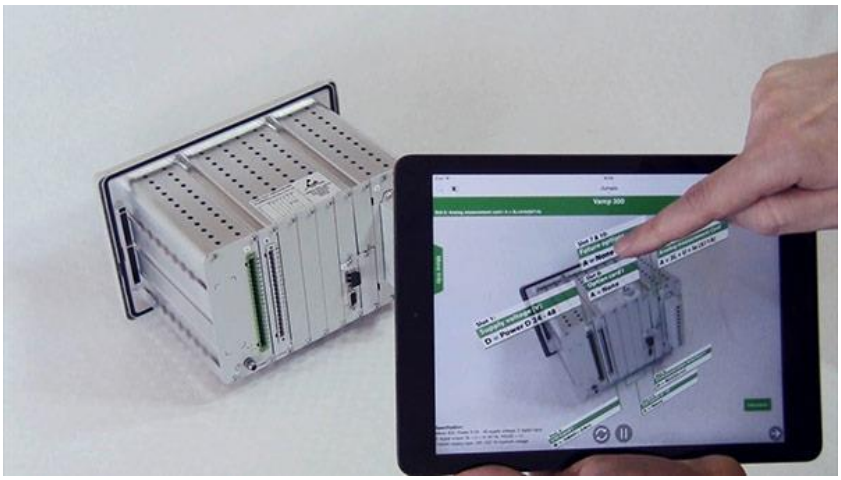

Figure 12: An example of Direct Manipulation of AR interfaces. Image credits 3D Studio Bloomberg

Touch-based interaction can be implemented following two different paradigms. The first is what interaction theorists call “Direct Manipulation” [17]: the user manipulates the virtual object and the interface directly using hand gestures. The virtual object or widget has properties similar to real-world objects so that the user can use her understanding of the physical world to manipulate the virtual imagery. Examples of this type of interaction are tapping on a button, swiping to rotate a 3D object or pinching the screen to zoom in. In AR user interfaces, the interaction is, in practice, very similar to 2D touch-based interfaces, with the difference that often content is not attached to the screen, but is rendered spatially at a precise point in the environment. In these cases, the user touches the visible area of the virtual object on the screen, trying to manipulate the object in space. Because of the similarity between spatially registered 3D content (which has a designated position and orientation in space) and manipulable virtual 3D content (which can be browsed and explored by direct manipulation), users often try to touch and manipulate spatially registered AR content in an attempt to explore it, instead of moving the point of view around the object itself.

Guideline: in cases where the user needs to explore a 3D model, enable for direct manipulation of 3D objects through touch interfaces.

The second implementation paradigm for touch-based interfaces is more indirect: external devices allow interaction with the interface through arm- or belt-worn surfaces connected to the presentation device. Buttons or touch surfaces sense the user's inputs and sends them to the presentation device that produces the expected effect on the interface. These devices can, in cases, provide a complementary information display.

Figure 13: Smart glasses connected to an external interaction device. Image credits Epson Inc.

This technique can be easy to implement if the presentation device can be connected to external devices, but it is not recommended for tasks during which body movements are spatially restricted.

Stare/gaze Tracking

Recently, stare/gaze input mechanisms have been implemented in some AR systems. These techniques allow to track the user's head or eye movements in order to interpret the interaction intention and act accordingly. In case of head tracking, a small pointer is displayed in the center of the view. The user stares at interactive elements of the virtual scene – overlays or UI components – and after a brief confirmation time, the input is accepted as intention. This technique, called "stare navigation," is best used for simple operations and manipulations as it can emulate only a single input – the equivalent of a mouse click.

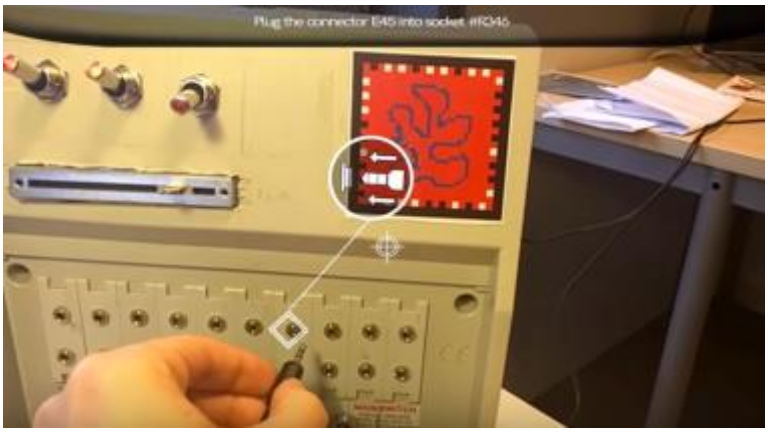

Figure 14: A stare/gaze navigation interface. The FOV-centered cross serves as interaction pointer.

A more complex interaction can be achieved using gaze-tracking technologies: micro cameras mounted in the near eye space track the pupil's position, and consequently, the point in the FOV towards which the user's attention is directed. Using the user's focus of attention, it is possible to detect what the intention is and perform an action without the user actively engaging with the system itself [18], or to stop displaying distracting information, thereby reducing visual clutter and saving power.

Multimodal Interaction

An optimal solution for interaction with AR systems may be to combine multiple techniques in order to mitigate the limitations of each one. Multimodal interaction can, ideally, provide the best user experience as long as it is correctly designed. Current technologies allow, for example, to easily combine gesture and speech interaction in order to grant the flexibility and intuitiveness of gesture-based input as well as the hands-free experience of verbal commands. In practice, however, there are limitations that can affect implementations of multimodal interaction techniques. Firstly, the more that input devices are combined, the more processing power is needed to detect all the different inputs, drastically reducing battery life – which is already short for wearable devices. Secondly, when more than one technique is involved during the interaction, achieving consistency and avoiding confusion requires extensive design knowledge and experience.

Once these technological limitations are overcome, multimodal input techniques facilitate the creation of more natural and intuitive interaction techniques, allowing users to focus on the task at hand, without having to be aware of how the underlying technology works.

Matching the Use Case with UX Design Priorities

At the beginning of this report, we listed some commonly known and frequently adopted use cases used as reference for AR implementations. The description of the use cases provides insights into the relevant characteristics that can influence AR system design. In this section, we match these characteristics with the considerations taken into account during the various sections of this report in order to provide an AR design solution to the three use cases examined.

The solutions proposed are not intended to be optimal. This section attempts to showcase how the guidelines and the discussions in the previous sections can be applied to real-life use cases. In addition, these solutions are not meant to be sufficient for every use case that matches the description. In fact there can be a number of external factors that influence the effectiveness of the proposed system. In these cases, the designer should follow the same critical mental process adopted in this report to analyze the characteristics of these external factors and determine how they impact the AR solution.

Warehouse Picking

From the description of this use case, it seems immediately obvious that users need to be mobile and easily roam around indoor facilities. Workers in warehouse facilities need to collect and deliver goods in different locations inside the building and move around on foot or with small vehicles. Consequently, a mobile solution is needed, thereby excluding projection-based stationary systems.

Another important consideration concerns the type of content that is going to be displayed. The essential types of information that need to be delivered by an AR application for warehouse support in order to support the successful fulfillment of basic tasks are:

- Checklists with the operations to perform or objects of interest (including the relevant information about their status)
- Textual descriptions of the operations, descriptions of objects, and notifications of status updates
- A simple navigation mechanism in the form of arrows or maps

This implies that 3D content and stereoscopy are not necessary for presenting this kind of

information, leaving space for easier-to-design and more computationally efficient 2D content. Monocular head-mounted displays can, then, be preferred to other types of display devices as they provide the hands-free experience always preferable for manual workers, but at the same time do not occupy the user's FOV.

Connectivity is a crucial component of the technology that influences the effectiveness of an AR solution for warehouse picking. For this scenario, Internet connectivity needs to be guaranteed at all times for instant order updating. Wi-Fi-enabled devices are usually a good fit as these tasks are often performed indoors. In addition, it is recommended to adopt devices that support the most recent standards for Bluetooth connection protocols. Bluetooth is used for indoor positioning supported by low-energy beacons installed in the facility, which are triangulated by the device to calculate the exact location.

As for content design, user-centric textual information and 2D content are sufficient to implement most of the functions required. Given the simplicity of this type of content, touch-enabled external devices or speech recognition – for low noise-level warehouses – are easy-to-implement solutions that work very effectively for 2D content.

Assembling a New Product

The assembly/disassembly of a mechanical or electronic component requires the operator to be able to picture the final configuration of the component and be aware of what operation is needed to reach that configuration and on what objects. This implies a high grade of spatial thinking that current paper manuals fail to provide through 2D schemes. For this reason, it is highly recommended to display 3D models of the initial and final configuration of the parts and animations to demonstrate how to fit the parts together or take them apart. In these cases, binocular smart glasses of handheld devices are recommended, as they are able to superimpose realistic 3D content on real objects. In addition, binocular HMDs can also render stereoscopic 3D content, further helping the visualization of spatial constraints in object placement, and they enable a hands-free experience, important during assembly operations that require both hands.

There are a number of use cases in which 3D spatial representation is not needed, as, for example, in welding operations. In this case it is far more important to precisely signal the welding location and a few other details like welder machine temperature or welding time. In this case, stationary projection-based AR systems are better solutions.

Projectors do not require the user to wear or hold a device, which would be uncomfortable

considering the amount of equipment already worn by a welder. Furthermore, projectors are less sensitive to the light condition variability caused by welding sparks.

As the content is mostly related to the object of interest and its components, object-centric overlays are usually implemented for supporting assembly: labels and images should be placed around the objects of interest while animated 3D models are superimposed directly onto objects, highlighting differences in configuration. Depth-sensing cameras are often used for these applications in order to improve the spatial precision of overlaid 3D models. In addition, by using depth maps, it is possible, in some cases, to automatically detect assembly errors or document operation compliance.

Maintenance and Repair Operations (MRO)

Many of the considerations for the assembly use case in the previous section are valid also for maintenance and repair operations. In fact, many MRO tasks include assembly and disassembly of physical objects. Consequently, spatially positioned overlays and animated 3D models are important for operators to build a correct mental model of the task and the objects of interest.

One key difference with the assembly use case lies in the extreme mobility required for MRO tasks. These tasks are mostly performed around facilities or even in remote sites. Consequently, stationary solutions are not convenient, while handheld and head-worn devices are often the recommended options. Smart glasses are, in general, a preferred device choice, allowing users to visualize while still being able to perform two-handed operations. However, handheld devices, such as tablets, have also proven to be effective and more cost-efficient solutions in less-complex scenarios that do not require the user to constantly pick up and put down the device.

Secondly, unlike in the object assembly scenario, the demand for network connectivity is much higher for MRO. These tasks often make use of Internet connections for a number of operations:

- Real-time data readings from industrial sensor networks, needed for equipment monitoring and fault diagnosis
- Status updates about concurrent maintenance operations or urgent interventions needed
- Expert remote assistance
- Job fulfillment documentation

Lastly, MRO may be performed in remote locations and highly variable environmental conditions. This requires the AR-enabled device to be durable and have a long battery life. Moreover, display technologies and interaction techniques need to be robust and adapt easily to diverse noise and lighting conditions.

Conclusions

This report provides general considerations and guidelines for AR experience design in industrial settings. Its goal is to address a broad spectrum of topics for audiences with a range of familiarity with Augmented Reality.

It explains decisions to be made throughout the design process and how choices impact the design of the AR experience. It serves as an introductory overview and may be useful for supporting discussions among people with diverse exposure to the field of Augmented Reality.

That said, we recommend keeping an open mind. Some of the topics introduced are very broad and complex and general guidelines cannot be applied, as the decision process varies according to the situation at hand. This should not stop the reader from reflecting upon the thought processes behind these guidelines and modifying them for their particular cases. More importantly, the structure of this report provides a guide for the decision-making pipeline and demonstrates how choices impact the design of the final experiences. We expect that the themes and guidelines will also need to be adapted as new devices are introduced and some obstacles to introduction decline in the future.

Bibliography

[1] J. M. Schraagen, S. F. Chipman, and V. L. Shalin, *Cognitive task analysis*: Psychology Press, 2000.

[2] N. A. Stanton, "Hierarchical task analysis: Developments, applications, and extensions," *Applied ergonomics,* vol. 37, pp. 55-79, 2006.

[3] M. R. Marner, R. T. Smith, J. Walsh, and B. H. Thomas, "Spatial User Interfaces for Large-Scale Projector-Based Augmented Reality," *Computer Graphics and Applications, IEEE,* vol. 34, pp. 74-82, 2014.

[4] M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas, "Spatial user interfaces for large-scale projector-based augmented reality," *Computer Graphics and Applications, IEEE,* vol. 34, pp. 74-82, 2014.

[5] S. R. Ellis, A. Wolfram, and B. D. Adelstein, "Three dimensional tracking in augmented environments: user performance trade-offs between system latency and update rate," in *Proceedings of the Human Factors and Ergonomics Society annual meeting*, 2002, pp. 2149-2153.

[6] D. Drascic and P. Milgram, "Perceptual issues in augmented reality," in *Electronic Imaging: Science & Technology*, 1996, pp. 123-134.

[7] C. Ware, *Information visualization: perception for design*: Elsevier, 2012.

[8] J. Paillard, "Motor and representational framing of space," *Brain and space,* pp. 163-182, 1991.

[9] S. J. Henderson, *Augmented reality interfaces for procedural tasks*: Columbia University, 2011.

[10] N. F. Polys, D. A. Bowman, and C. North, "The role of depth and gestalt cues in information-rich virtual environments," *International journal of human-computer studies,* vol. 69, pp. 30-51, 2011.

[11] B. Bell, S. Feiner, and T. Höllerer, "View management for virtual and augmented reality," in *Proceedings of the 14th annual ACM symposium on User interface software and technology*, 2001, pp. 101-110.

[12] M. Tatzgern, D. Kalkofen, R. Grasset, and D. Schmalstieg, "Hedgehog labeling: View management techniques for external labels in 3D space," in *Virtual Reality (VR), 2014 iEEE*, 2014, pp. 27-32.

[13] M. Sukan, C. Elvezio, O. Oda, S. Feiner, and B. Tversky, "ParaFrustum: visualization techniques for guiding a user to a constrained set of viewing positions and orientations," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, pp. 331-340.

[14] M. Fiorentino, A. E. Uva, and G. Monno, "Product manufacturing information management in interactive augmented technical drawings," in *ASME 2011 World Conference on Innovative Virtual Reality*, 2011, pp. 113-122.

[15] P. Tiefenbacher, T. Gehrlich, and G. Rigoll, "Impact of annotation dimensionality under variable task complexity in remote guidance," in *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, 2015, pp. 189-190.

[16] A. D. Wilson, "PlayAnywhere: a compact interactive tabletop projection-vision system," in *Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005, pp. 83-92.

[17] B. Shneiderman, "1.1 direct manipulation: a step beyond programming languages," *Sparks of innovation in human-computer interaction,* vol. 17, p. 1993, 1993.

[18] J. Orlosky, T. Toyama, K. Kiyokawa, and D. Sonntag, "ModulAR: Eye-controlled Vision Augmentations for Head Mounted Displays," *Visualization and Computer Graphics, IEEE Transactions on,* vol. 21, pp. 1259-1268, 2015.