# Investigating the Use of AR Glasses for Content Annotation on Mobile Devices

Francesco Di Gioia, Eugenie Brasier, Emmanuel Pietriga, Caroline Appert

# Investigating the Use of AR Glasses for Content Annotation on Mobile Devices

FRANCESCO DI GIOIA, EUGÉNIE BRASIER, EMMANUEL PIETRIGA, and CAROLINE APPERT, Université Paris-Saclay, CNRS, Inria, LISN, France

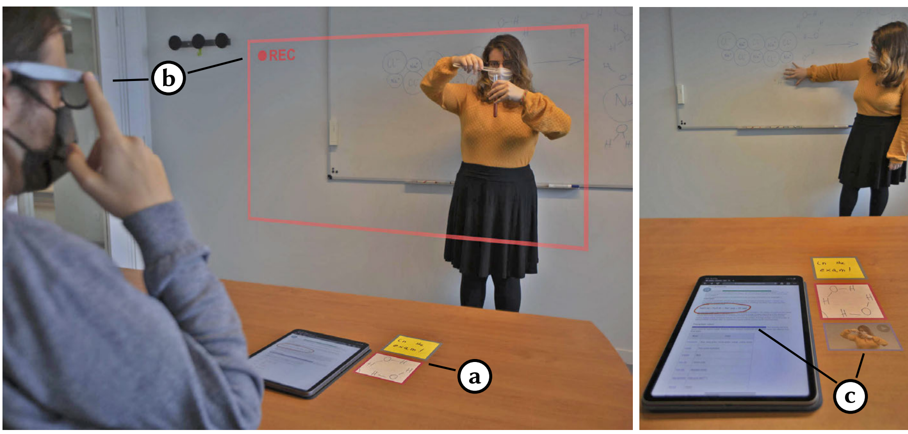



Fig. 1. An artist's impression of a student annotating content on a tablet with the help of smartglasses during a Chemistry course. The tablet displays course material related to the experiment that the teacher is performing live. The student's smartglasses display annotations made to that course material around the tablet (a). The student also uses the smartglasses to capture the experiment as a video that he then inserts as an annotation (b), that eventually gets displayed with links to the relevant part of the course material (c).

Mobile devices such as smartphones and tablets have limited display size and input capabilities that make a variety of tasks challenging. Coupling the mobile device with Augmented Reality eyewear such as smartglasses can help address some of these challenges. In the specific context of digital content annotation tasks, this combination has the potential to enhance the user experience on two fronts. First, annotations can be offloaded into the air around the mobile device, freeing precious screen real-estate. Second, as smartglasses often come equipped with a variety of sensors including a camera, users can annotate documents with pictures or videos of their environment, captured on the spot, hands-free, and from the wearer's perspective. We present *AnnotAR*, a prototype that we use as a research probe to assess the viability of this approach to digital content annotation. We use *AnnotAR* to gather users' preliminary feedback in a laboratory setting, and to showcase how it could support real-world use cases.

CCS Concepts: • **Human-centered computing** → **Mixed / augmented reality**.

Additional Key Words and Phrases: augmented reality; mobile device; annotation

**ACM Reference Format:**
Francesco Di Gioia, Eugénie Brasier, Emmanuel Pietriga, and Caroline Appert. 2022. Investigating the Use of AR Glasses for Content Annotation on Mobile Devices. In . ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3567727



*ISS '22, November 20–23, 2022, Wellington, New Zealand*
2022. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/3567727

## 1 INTRODUCTION

Mobile devices such as smartphones and tablets can be used to browse the Web, to read documents, to view image collections, to navigate maps, or to watch videos. Such content has become easily accessible, not only in mobile situations but in other contexts as well, such as at home or in the workplace, where it is sometimes more efficient or more comfortable to use a lightweight touch-enabled device sitting nearby rather than walk to a desktop workstation or fetch a laptop.

A mobile device will often be sufficient when the activity is limited to consuming content or performing lightweight edits to a document. Meanwhile, recent improvements – support for multitasking or advanced note-taking frameworks (*e.g.*, [8]) – aim at broadening the range of activities that users can perform on their mobile devices. Yet, limitations in terms of both screen size and input capabilities can still make more elaborate content editing and annotation activities frustrating, leading users to delay their task until they have access to a more suitable platform.

Content annotation in particular is often proving challenging on a mobile device. Screen real-estate is at a premium, and annotations will be competing for display space with the very content that users are trying to annotate. Taking text documents as an example, classic margin notes – as found in Word or PDF documents – require space on one side of the document. But the lack of screen space for displaying annotations is not the only issue. Depending on the mobile device’s input capabilities, the annotation process itself can be tedious. While a pen+touch interface will help highlight text or a region on a map and then use ink as input, small devices that only support touch input and text entry using a soft-keyboard will often make the annotation process frustrating.

In this paper, we investigate how Augmented Reality (AR) eyewear can help streamline annotation creation and consumption on mobile devices. This particular cross-device setup enables offloading annotations into the air around the handheld device. It also enables creating annotations using pictures and videos captured on the spot, from the wearer’s perspective, hands-free (see Figure 1). We present *AnnotAR*, a research prototype that we use as a probe to assess how effective this approach to supporting annotation tasks can be. Our prototype enables annotating a variety of contents that users consume on mobile devices: Web documents, maps, and videos. Users can create annotations of multiple types: not only text or freeform ink, but pictures or videos as well, captured with the eyewear’s embedded camera. We use *AnnotAR* to conduct a small-scale study providing preliminary feedback on the approach in a laboratory setting, and to run more realistic scenarios that illustrate how users could benefit from the approach in a real-world setting.

Our contributions are:

- an AR-enhanced approach to digital content annotation tasks performed on mobile devices;
- a proof-of concept prototype that implements this approach;
- an evaluation of this approach through both empirical feedback and a showcase of the user experience with a set of actual, real-world scenarios.

## 2 RELATED WORK

AR has been primarily studied as a means to annotate physical content, leveraging AR display to register digital information with objects and facilities in the user’s environment [5, 11, 46], including paper documents [38]. We take a different approach, in which AR is used to annotate *digital* content, displayed on a mobile device, with a variety of types of annotations. Our contribution builds upon previous work about both digital content annotation and cross-device AR systems.

### 2.1 Annotating Digital Content

Most annotation systems are designed for desktop workstations. Some applications such as Adobe Acrobat and GoodReader enable annotating PDF documents on tablets and smartphones. But these applications mostly replicate the annotation workflow of the desktop, which is typically limited to highlighting text and inserting margin notes. Some annotation systems also support digital pen input to draw marks, building on the concept of the *Active Reading Machine* introduced in the XLibris project [16, 17, 42]. Annotations are usually displayed as overlay marks or as free-floating, Post-It-like widgets.

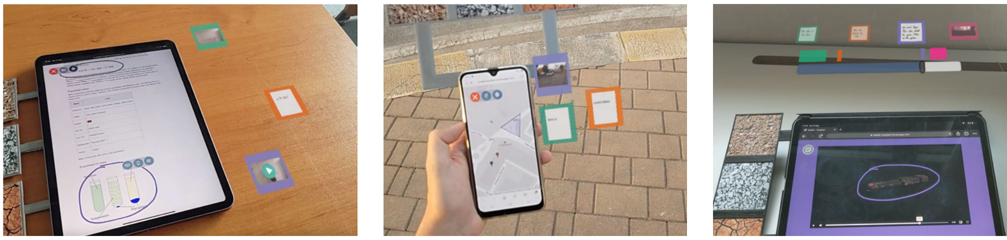
While some annotation systems have been designed for large surfaces such as tabletops [31], they often run on devices where screen real-estate is an issue – as is also frequently the case with printed paper and an analog pen. Several strategies exist to address the lack of display space for annotations. For example, the DIZI system [1] pops up a magnifying lens when users start annotating in order to enlarge the space between lines. While this facilitates input, annotations will still be small and difficult to read when not magnified. Other approaches actually alter the document to make space for annotations. For example, TextTearing [47] enables users to create space for annotations between two paragraphs, effectively shifting down part of the page's content. SpaceInk [41] pushes this idea further. Leveraging the reflowable property of digital document formats such as HTML, it lets users make space for in-context annotations between words, lines or paragraphs. Such annotations are tightly integrated with the corresponding content, but they also consume screen space, change the document itself, and decrease the amount of content of interest that can fit on screen.

LiquidText [44] takes a different approach by clearly separating the content of interest from the annotations. The device's screen is divided into two areas. The left side shows the main document, while the right side consists of a workspace for managing annotations. This clear spatial separation means that the primary content and the annotations are not directly competing for display space and do not visually interfere with one another, as they do when tightly intermingled. But this separation also reduces the display space available for the primary content, as an entire monolithic region has to be dedicated to annotations.

No matter the level of integration or separation between content and annotations, there is some tension between them in terms of screen real-estate use. This tension is particularly prevalent on mobile devices, which have a very limited display capacity. On the desktop or any other stationary context, display capacity can be increased by adding more screens or using larger ones with a higher resolution. This is hardly possible in a mobile context, in which lightweight devices that can be operated with a single hand tend to be favored. Instead of increasing the physical display area, our approach uses the air around the device to display annotations, effectively freeing space on the mobile device for the primary content.

### 2.2 Cross-Device AR Systems

The combination of a mobile device with AR eyewear opens up a large design space of promising user interfaces, as one device can compensate for the weaknesses of the other device. The BISHARE design space [49] organizes smartphone + AR eyewear combinations into two broad categories: *HMD-centric* and *Phone-centric*. In the first category, the content of primary interest is displayed in AR, and the mobile device often serves as a controller for manipulating that AR content as in, *e.g.*, DualCAD [32], TrackCap [33] or the work of Budhiraja *et al.*[10]. Conversely, in the second category, the mobile device displays the content of primary interest. AR is mainly used to enhance the presentation and interactive manipulation of that content. Our approach belongs to the latter



a) b) c)

Fig. 2. Different types of digital content annotated with *AnnotAR* using a HoloLens 2: a) Web document, b) map, c) video.

category, investigating the use of AR for a particular type of task that, to the best of our knowledge, has not been considered so far: digital content annotation.

AR can augment phone-centric applications in several ways. VESAD [35] extends a phone's screen with an AR plane that is spatially aligned with the phone. The direct application of a VESAD is to provide a wider view on large information spaces in order to show more information, that would otherwise fall off the phone's screen. Although recent empirical results suggest that having to reconcile the two displays adds to the users' cognitive load [15], a VESAD can still prove more efficient than a phone alone for tasks such as, *e.g.*, manual classification [35].

The additional display space around the phone can also be used for other purposes than enlarging a single, continuous scene. It can show information that is complementary to that displayed on the phone but of a different nature, as in WatchThru [45] and mobile true-3D displays [43]. MultiFi [18] redesigns a series of UI widgets to make them run across a mobile phone and an AR headset in order to take advantage of these devices' respective fidelities. AR-enhanced widgets [7] are traditional UI components designed to be offloaded to the air in order to free space for the content of interest on the phone.

In this paper, we contribute to the exploration of this large mobile device+AR design space, studying how smartglasses can be leveraged when annotating digital content on a mobile device. To this end, we developed a prototype system that we use as a research probe to evaluate the potential benefits of smartglasses, including both sensing and AR display capabilities.

## 3 ANNOTAR PROTOTYPE

*AnnotAR* is a proof-of-concept AR-enhanced mobile annotation system. It enables annotating a variety of content types that users frequently consume on mobile devices (Figure 2), including: documents (such as Web pages) as well as maps and videos. *AnnotAR* runs on AR head-mounted displays (ARHMD) such as the Microsoft HoloLens 2.

*AnnotAR* lets users create annotations in multiple ways. These include many classic techniques such as highlighting, freeform inking and Post-It notes containing either typed text or handwriting. Typing and handwriting often prove tedious on a handheld device, however. Annotating with recorded voice notes gives users an alternative [48], but voice carries a limited amount of information and may not be socially comfortable in all situations.

*AnnotAR* provides users with additional means of capturing annotations, leveraging the ARHMD's embedded camera – a feature already found in many commercial smartglasses, including the most lightweight ones. The camera lets users annotate digital content with images and videos captured live. It is also possible to annotate using media captured previously and stored in the mobile device's photo gallery. This new type of annotation can be particularly useful in scenarios where the stimuli

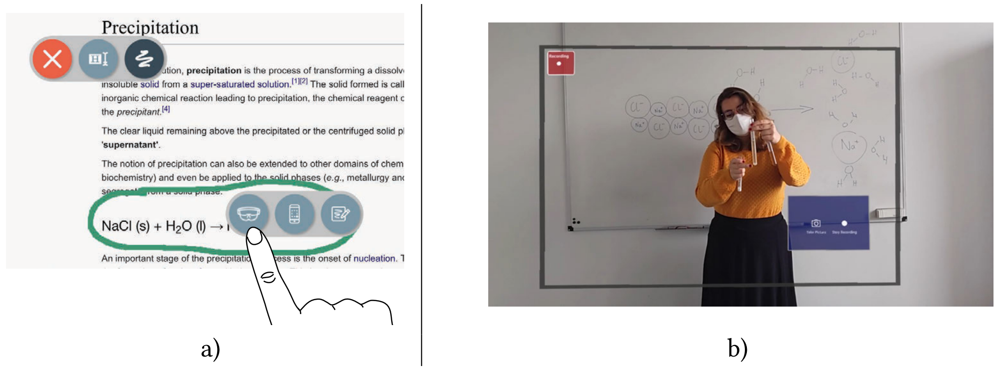



Fig. 3. Annotating part of the course material (formula of a chemical reaction) with a video clip of the professor actually performing the experiment live. a) Using the freeform tool to circle the formula on the mobile device, and then selecting the ARHMD's camera as the annotation input source. b) Controlling video capture with buttons displayed in AR (take a snapshot, start/stop recording).

for annotating are mostly visual or come from the environment, such as the chemistry experiment performed by the teacher in Figure 1, described in detail in Section 5.1.

Annotations created with *AnnotAR* are displayed in AR using the ARHMD, as if they were floating in the air around the mobile device. This serves two purposes:

- optimize the mobile device's display for the primary content by offloading annotations around it;
- keep annotations clearly distinguishable from the original content, as advocated by previous studies which reported this to be an important feature [30, 36].

We now describe how users can input, display and interact with annotations in this mobile+AR configuration.

### 3.1 Input

As with most annotation systems, the workflow consists of two main steps: defining the scope of the annotation, and actually inputting the annotation.

*3.1.1 Scope Definition.* A button on the mobile device's display lets users enter annotation mode. The button unfolds into a menu from which users can select tools to define the annotation's scope: a freeform inking tool, and a selection tool (Figure 3). The first tool lets users draw anywhere over the primary content with digital ink and then associate an annotation with that ink. The second tool lets users highlight any *Media Unit*. The definition of a Media Unit depends on the content type considered. This can for instance be a text span in a document; a time interval in a video; a marker or any geographical entity – such as a building – on a map.

*3.1.2 Annotation Input.* Once the scope is set in the content, a menu pops up, featuring three different tools.

The first tool creates a regular annotation in the form of a *Post-It Note*. Users can either type text with the mobile device's soft keyboard, or draw marks using their finger or a stylus.

The second tool lets users annotate content with pictures and videos from their mobile device. These can be either captured live with the mobile's camera, or imported from its photo gallery. This can be particularly useful for small devices, on which the single-view model is prevalent. This indeed enables users to easily annotate with content coming from different applications, taking screenshots of those if need be. For instance, one can annotate the name of a city in a text-based document with a screenshot of this city's map or a picture of its skyline. It also supports the sort of side-by-side comparisons that active readers often make [34, 36].

Capturing live pictures and videos with a mobile device as described above requires holding it in a way that might not always be comfortable, or that might be socially awkward. The third tool, illustrated in Figure 3, provides an alternative which consists of using the smartglasses as an input source instead of the handheld device to perform the live capture (either picture or video). This feature, enabled by the ARHMD's camera in our prototype, is particularly interesting when annotating content with a snippet from the real world that can be captured hands-free, simply looking around.

### 3.2 Output

The ARHMD is in charge of displaying annotations in the air, as if they were floating around the mobile device. The mobile device's screen can then be fully dedicated to the primary content. But having the primary content and annotations displayed in two separate spaces creates an indirection between them. We use the following strategy to make annotations easy both to access and to relate to their scope.

*3.2.1 Scope-Annotation Association.* Systems that support the annotation of Web documents must cope with variations in their layout, which can change depending on the display configuration [9, 12, 37]. Similarly, *AnnotAR* dynamically updates which annotations are displayed in the air depending on the current viewport configuration so as to show a subset of them, aiming to show the most relevant ones. Annotation visibility and spatial alignment with the scope is not sufficient to let users easily associate them visually. We originally considered adding explicit visual encodings to relate annotations to their respective scopes, such as drawing curved lines between them. We abandoned this idea for two reasons. First, it quickly adds visual clutter, making links difficult to follow and hindering the readability of the primary content. Second, technical limitations with current hardware sometimes result in slight visual misalignments (in mobile situations in particular). Such misalignments are disturbing, especially with visual encodings that strongly rely on position. We eventually opted for an indirect encoding of the association using color hue. When defining a new scope, *AnnotAR* automatically assigns a color to it. The color is picked from the *Dark2* Color Brewer palette [20], which is optimized for encoding categorical data, and is resistant to color-blind confusions. The number of different colors that users can easily differentiate in a categorical palette is typically limited to less than a dozen [20], but this should be sufficient in most situations: the count of annotations displayed simultaneously should rarely exceed this number, given that *AnnotAR* only shows the subset of annotations associated with scopes currently visible in the handheld's viewport. In the rare cases where it does exceed this limit, spatial alignment between annotations and their scope will provide an additional visual cue that will help disambiguate associations involving the same color, as they will be far away from one another with annotations using different colors in-between.

*3.2.2 Annotation Layout.* Based on findings from previous studies about the layout of AR content relative to handheld devices [7, 35], we place annotations on the side corresponding to the user's dominant hand when the device is held in portrait mode (*e.g.*, text documents, maps on smartphones, see Figure 4-a); and above the device's top edge in landscape mode (*e.g.*, videos, Figure 4-b). In-air annotations are displayed on the side opposite to the non-dominant hand – which typically holds the device – thus limiting cases where users would have to cross their arms to access annotations.

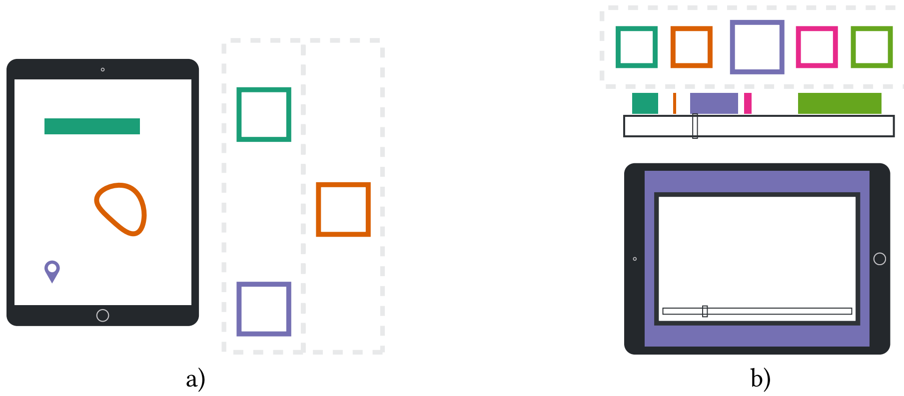



Fig. 4. Annotation display spaces for a right-handed user. The annotation display space can be a) on the side of the dominant hand in portrait mode; b) above the device's top edge, together with the timeline when showing a video.

In this proof-of-concept prototype, annotations are displayed as fixed-size square-shaped thumbnails that users can enlarge by performing in-air tap gestures. More elaborate thumbnail generation strategies, such as content-aware cropping or content simplification & rescaling, could be investigated in future iterations.

The main problem, however, is about keeping annotations properly anchored to their scope. Systems that support the annotation of Web documents [9, 12, 16, 37, 41] – whose layout can vary depending on the display configuration and on the user's preferences – have already faced this problem. *AnnotAR* implements the following behavior: any navigation action on the handheld device fires a notification to the AR headset, which then updates the layout of annotations in the air based on the new viewport configuration.

For text documents and maps, which are typically displayed in portrait mode, annotations are displayed on the side of the dominant hand, making the assumption that the device is held with the non-dominant hand (Figure 2-a and 2-b). The vertical position of annotations is a function of the vertical position of their scope on the handheld device. Annotations are ordered according to the scope's vertical position, color helping disambiguate between scopes that are close to one another. As mentioned earlier, by default, only annotations that correspond to scopes visible in the handheld's current viewport are displayed.

As mentioned earlier, *AnnotAR* also enables annotating videos. This specific type of content is handled a bit differently. First, an annotation in a video can have a spatial scope (a region in the frame), but it will often have a temporal scope as well (typically a time interval, but possibly a single frame). Second, as users frequently watch videos with their device held in landscape format, *AnnotAR* rather displays annotations in the air above the handheld device (Figure 2-c) rather than on its side, to optimise layout. This annotation display space is used not only to show preview thumbnails of annotations that are temporally close to the current frame, but also to display a timeline decorated with smaller markers for all annotations in the video (Figure 4-b). The annotation thumbnails get updated as the video plays. When the current frame falls in an annotation's temporal scope, its thumbnail gets magnified and the video's margins get painted with that annotation's color.

### 3.3 Implementation

As full-featured AR smartglasses are not available on the market yet, we use a Microsoft HoloLens 2 ARHMD. We couple it with either a Samsung Galaxy A50 or an iPad Pro 2020, but *AnnotAR* can run on a broad range of mobile devices since we only use Web-standards-compliant technologies.

*AnnotAR* is based on two client applications and one server. The client application for the handheld device is implemented in HTML+Javascript. The following third-party libraries are used to handle different content types and to process input: Leaflet.js for maps, video.js for videos and jSketch[1] for freeform marks. The client application running on the headset is built with the Unity game engine. It is developed with C# and the Microsoft Mixed Reality Toolkit (MRTK). The server enables communication between these two client applications. It is implemented as a Node.js application deployed with Heroku. Messages are serialized using the JSON format and exchanged using Socket.IO (binary data such as images are encoded using a Base64 scheme).

Navigation actions on the mobile device fire notifications to the ARHMD, which then updates the layout of annotations in the air based on the new viewport configuration.

## 4 PRELIMINARY USER STUDY

We used the *AnnotAR* prototype as a probe to gather preliminary empirical data about people's ability to use an interactive system coupling a mobile device and smartglasses for digital content annotation. We conducted a small-scale user study, aiming to assess the viability of such an approach, seeking to gain insights about two fundamental aspects: 1) participants' ability to relate annotations with their scope in the primary content, given that the former are displayed in the air and only implicitly linked to the latter; and 2) their ability to use both devices (handheld and eyewear) together to capture annotations and how seamless the process can be. Beyond assessing participants' ability to perform those tasks, our goal was to collect qualitative feedback about the approach overall and about the prototype system as well to inform future design iterations.

We considered both the real-time creation of annotations and the reading of already-created annotations. In a first phase, participants had to run specific scenarios that we scripted in order to make them *create* different types of annotations on the spot, more particularly those involving the smartglasses camera as an input source. In a second phase, participants had to relate an annotation to its scope and *vice versa* in order to observe how difficult it might be to relate annotations in the air with the corresponding content snippets on the mobile device. In order to collect user feedback, we instructed participants to *think aloud* all along the experiment. We also asked them to assess the usability of the different features using the System Usability Scale (SUS) questionnaire. [2]

### 4.1 Participants

Six participants (5 men and 1 woman, aged 32 ± 10) volunteered for the experiment. Participants were asked about their prior experience with Head-Mounted Displays (Augmented or Virtual Reality headsets). One participant uses AR on a regular basis, while the five others are familiar with the technology but have used it only on a few occasions.

### 4.2 Apparatus

*AnnotAR* ran on a HoloLens 2 combined with either a Samsung Galaxy A50 smartphone or an iPad Pro 2020 tablet, depending on the task. To avoid potential uncontrolled effects due to inaccuracies when tracking the mobile device's position, participants were seated with that device positioned

[1]https://luis.leiva.name/jsketch/

[2]Study material available at https://ilda.saclay.inria.fr/AnnotAR

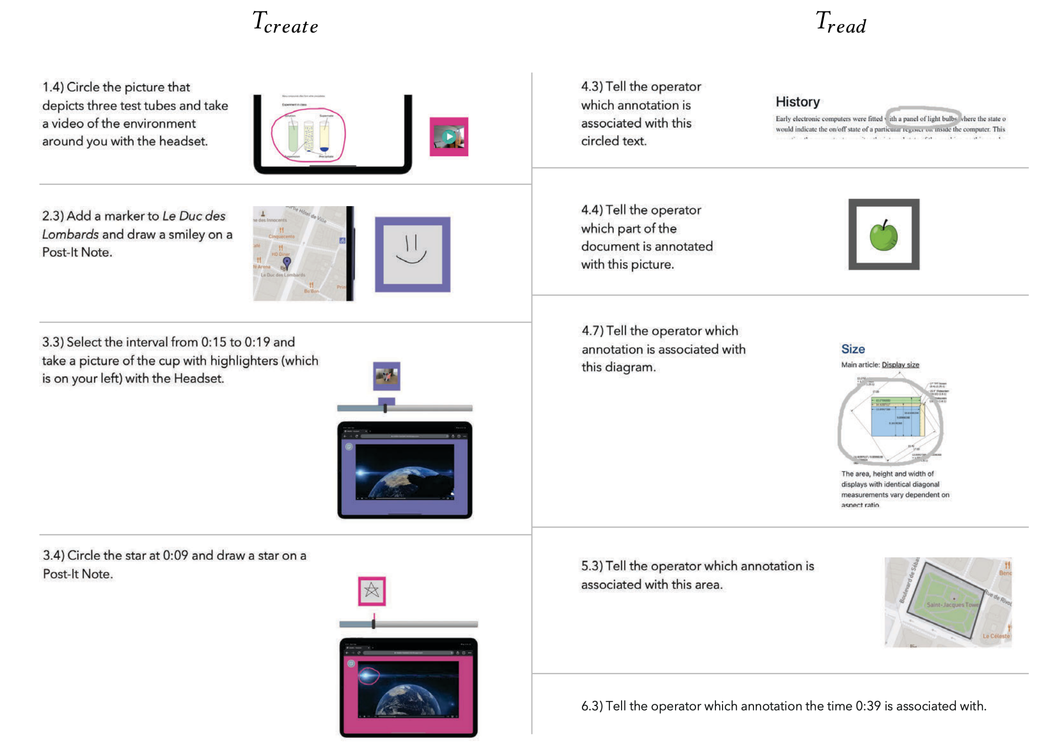



Fig. 5. Representative tasks from the study. Left column: four of the twelve $T_{create}$ tasks. Right column: five of the twenty $T_{read}$ tasks. Annotations were associated with either 1) text or an image in a document; 2) a marker or area on a map; 3) a specific time or interval in a video, sometimes coupled with a specific area in the picture.

on a stand on the table in front of them. The operator monitored and controlled progress through the task sequence thanks to an application running on their laptop.

### 4.3 Tasks

Participants had to perform two types of tasks which operationalize cases where users either 1) create annotations ($T_{create}$) or 2) read annotations that already exist ($T_{read}$). Figure 5 shows a representative sample of tasks. The tasks were printed on a sheet. In a $T_{create}$ task, participants were presented with some content on the mobile device without any annotation, and were asked to create an annotation. In a $T_{read}$ task, participants were presented with some content on the mobile device that had already been annotated with *AnnotAR*. They were asked to associate an annotation with the corresponding content snippet (scope) or *vice versa*.

### 4.4 Design and Procedure

Participants entered the experiment room and sat at a desktop workstation. They read & signed a consent form. Then the operator started with a demonstration of what can be done using three video clips, one per type of content (the companion video shows excerpts from these video demonstrations). The operator also distributed a tutorial sheet listing the different features of

*AnnotAR* and the main steps to use them. The experiment was divided into two phases: $T_{create}$, and then $T_{read}$.

*4.4.1 Phase 1: creating annotations.* Participants had to perform a series of $T_{create}$ tasks. As our goal was to focus on the AR-enhanced features (*i.e.*, annotations displayed in the air and annotation input involving the ARHMD), we scripted a series of 6 different tasks that participants walked through, involving *AnnotAR*'s ARHMD-camera feature for each of the three content types (text document, map, and video). The other three tasks were chosen so as to ensure a representative sample of combinations of scope types and annotation input technique over the experiment. Each task was replicated twice in a row, resulting in a series of 12 tasks. The first trial operationalizes a situation in which *AnnotAR* is used for the first time for a given annotation task. The second trial operationalizes the use of *AnnotAR* after having discovered that feature. Tasks were grouped by content type. They were always presented in the same order.

Each $T_{create}$ task started with participants reading instructions about how to complete the following task. Instructions consisted of a short text description of the task as well as a picture of what the system should look once the task was completed (see examples in Figure 5). Participants were free to refer to the tutorial sheet at any time.

At the end of Phase 1, participants had to fill-in two SUS questionnaires. The first questionnaire asked participants to evaluate feature *Inserting annotations capturing photos or videos with the Headset*; the second one did the same for feature *Inserting annotations from the Handheld Device or as a Post-It Note.*[3]

*4.4.2 Phase 2: reading annotations.* Participants had to perform a series of twenty $T_{read}$ tasks. A $T_{read}$ task consisted of identifying which annotation was associated with a given scope, or conversely. As in the first phase, the task series was always presented in the same order. It consisted of three blocks, one per content type. There were 8 $T_{read}$ tasks on text documents and 8 $T_{read}$ tasks on maps. The annotation was visible in the initial viewport for half of the trials. In the other half, participants had to navigate the document or map to find the annotation. Because of the temporal nature of videos, there were only 4 $T_{read}$ tasks performed on a video, each of which involved navigating through the video to complete it. In each block, the tasks involved the two types of scope (freeform or media unit). For each scope, there was one task in which users had to find an annotation from its scope, and another task where they had to find the scope of a given annotation. See examples in Figure 5.

A $T_{read}$ task started with participants reading printed instructions about how to complete the next task. When the task consisted of finding the annotation for a given scope, instructions consisted of a picture of the scope only, without any color coding (see, *e.g.*, Task 4.3 in Figure 5). Participants had to search for that scope and tell the operator what the associated annotation was. Annotations systematically took the form of pictures from the Dobble game[4] which, as stated in the rule book, are designed to be easy to name. When the task consisted of finding the scope of a given annotation, instructions consisted of a picture of the annotation only. Participants had to search for the associated scope and tell the operator. When performing tasks on the $Text$ content type, participants had to read the text that was circled or highlighted. For $Map$s, they had to read the associated label. For $Video$s, they had to tell the frame's timestamp.

At the end of each of Phase 2's three blocks, participants had to fill in an SUS questionnaire about feature *Reading annotations* for the type of content they had tested in the block ($Text$, $Map$ or $Video$).

[3] All 10 statements of the SUS questionnaire use the word *system*. We replaced that term with the term *feature* instead.
[4] https://www.dobblegame.com/en/homepage/

### 4.5 Results

The collected data consisted of both open remarks captured through a think-aloud protocol and numeric grades given in the SUS questionnaires. Participants typically read the tutorial sheet before starting the experiment and sometimes returned to it before using a feature for the first time. Participants successfully completed all tasks. They were overall very positive, with comments such as *"After executing it the first time, it is really simple."* (P1) or *"All easy and straightforward"* (P3). These preliminary results suggest that users are indeed able to use both devices in a seamless workflow when creating annotations, and that they do not have difficulties relating annotations to the corresponding content snippets.

Below we report participants' assessment of *AnnotAR*'s usability through their normalized scores to the different SUS questionnaires as well as the feedback they gave regarding possible improvements.

*4.5.1 Phase 1: creating annotations.* As mentioned above, we asked participants to rate their experience when creating annotations with either the ARHMD's camera or other means (*e.g.*, with a Post-It note or with some picture from their mobile device's gallery) *separately*. Although the SUS score was a bit lower for the ARHMD (87.1 ± 5.3) than for other means (92.9 ± 6.8), both scores were consistently very high across participants. Unsurprisingly, two participants (P1,P4) mentioned that the real issue is the ARHMD itself, as the technology is not quite mature yet. Its lower score might also come from a lack of feedforward and feedback when annotating with a picture captured using the ARHMD. Participants would have liked to see a frame delimiting the area to be captured, as featured when recording a video (P4,P5). They also commented on the Take picture button and confirmation message not being visible enough (P3,P5), suggesting that these widgets should be more central in the view.

*4.5.2 Phase 2: reading annotations.* The SUS scores about reading annotations on all three types of content were very high: 94.2 ± 8.3 ($Text$), 93.3 ± 3.7 ($Map$) and 93.8 ± 7 ($Video$). All participants immediately understood that the correspondence between content snippet (scope) and annotation is achieved implicitly with color coding. There were very few comments about the reading of annotations. All participants found it clear and easy. One participant (P5) would have liked the annotations to be better aligned with the corresponding snippet when there are not too many in the same view though, which suggests that more elaborate in-air annotation layout strategies should be investigated.

Participants suggested several improvements. Three participants discovered by themselves that they could tap an in-air annotation to enlarge it, commenting positively on that feature. But some participants would have liked more in-air interaction. In particular, P3 and P6 suggested displaying a line between an annotation and its scope when hovering an element (in the spirit of what MARVIS [27] does when brushing and linking between two visualizations). P5 would have liked to use an annotation thumbnail as a navigation cue to directly jump to the corresponding time in videos. Finally, P6 expected the video timeline in the air to enable them to interactively navigate the video. Three participants (P1,P2,P5) would have liked annotations whose scope is no longer in the map's viewport to remain visible, displayed in the air on top or below (P2) or as a cloud around the phone (P1). In-air annotations could indeed be an alternative or complement to techniques for indicating the presence of objects offscreen on small devices [4, 19].

## 5 ANNOTATING AWAY FROM THE WORKSTATION

Feedback from the study suggests that the approach is indeed viable, as participants were able to use the combination of AR headset and handheld device to capture new types of annotations.

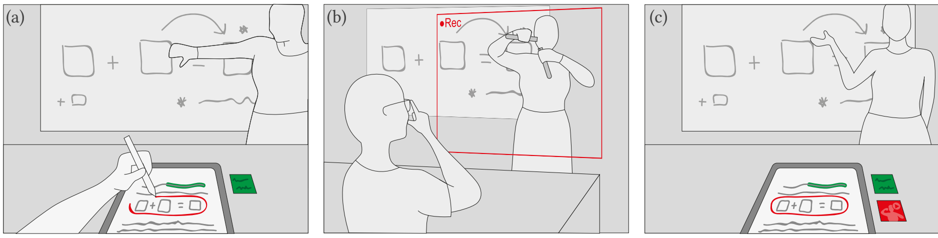



Fig. 6. $S_{text}$ scenario: a) Kim has already input a handwritten annotation, displayed as a green note to the right of their tablet, and are now circling a figure in the document to annotate it with a video of the teacher. b) Kim uses their smartglasses to capture a video of the teacher performing an experiment live. c) Once the recording has stopped, the video appears as a new annotation on the tablet's right side.

We now describe three scenarios that illustrate what sort of user experience the combination of a mobile device with smartglasses could enable. All three scenarios have actually been implemented with *AnnotAR* to demonstrate their technical feasibility.[5] While we had to use current hardware for their implementation, the scenarios are situated in the near future, where AR eyewear has become much more lightweight than current ARHMD devices, effectively becoming smartglasses that can realistically be adopted by users in their daily activities. The three scenarios involve both mobile and stationary contexts that can benefit from AR-enhanced annotation features.

## 5.1 Scenarios

*5.1.1 $S_{text}$ – Annotating a text document (Figure 6 and Figure 1).* Kim is attending a Chemistry class. They use their tablet to display the course material provided by the professor. It consists of a text document with pictures and diagrams. Kim takes their own notes, but also likes to annotate the digital material directly on their tablet: highlighting text, circling particular items that the professor is putting emphasis on.

In order to explain a specific chemical reaction, the professor draws a diagram on the whiteboard to explain how the different atoms and molecules get reconfigured during this reaction. Kim finds the diagram much easier to understand than the concise formula in the document. They would like to capture this diagram, that will definitely help remember when preparing for the exam. Kim circles the formula on the tablet, takes a picture of the whiteboard with their AR smartglasses and inserts it as an annotation.

The professor then performs a live demonstration of the experiment that students will have to reproduce in the next practical session. A schema in the course document describes the experiment, but it only gives a rough outline of the main steps (Figure 6). Kim uses their glasses to record a video of the professor actually performing the experiment. The video is then linked to the course material as an annotation to the schema.

In both cases, Kim only had to tap on the side of their glasses to record the video or take a picture of what they were seeing. This frees Kim from having to raise any handheld device to capture the image, streamlining the process of capturing such types of annotations. It also makes it more socially acceptable. In addition, having the annotations displayed "in the air" next to the course material has several benefits. 1) The extended display space means that annotations can be shown as previews, improving annotation-driven navigation in the document. 2) The annotations remain easily distinguishable from the main course material. 3) They do not take any space on the tablet itself, whose display remains entirely dedicated to the main material.

[5]The companion video to this article shows the scenarios running with the *AnnotAR* prototype.

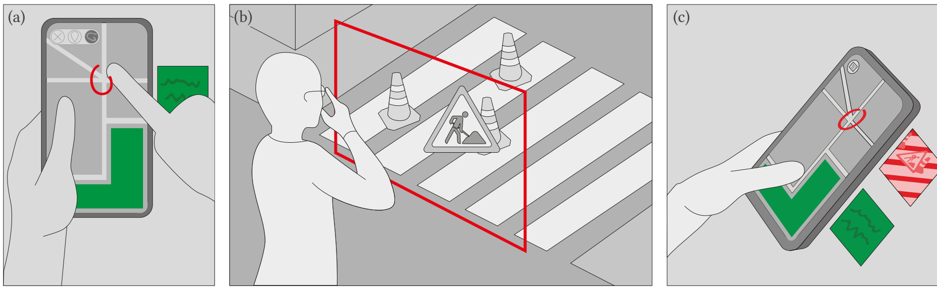



Fig. 7. $S_{map}$ scenario: a) Laura has already created a text annotation about a new restaurant in the green building, and she is now circling another area she wants to annotate (a road intersection); b) she then takes a picture of the intersection with her smartglasses. c) The picture appears as a new annotation, floating in the air on the side of her smartphone.

*5.1.2 $S_{map}$ – Annotating a map (Figure 7).* Laura regularly contributes to OpenStreetMap,[6] a collaborative project to create an open geographic database of the entire world. There is a lot of construction work going on near Laura’s workplace, and she wants to keep the database up to date as often as possible. She regularly walks in the area, with an OpenStreetMap editor running on her smartphone, comparing what is on the map and what she can see in the field to identify inconsistencies and make edits accordingly.

Laura notices that a new restaurant has opened on the ground floor of a recently-constructed building. The building is already on the map, but has no tags. She selects the building and adds a tag to it in the form of an annotation, filling in information such as its type (a café), its precise address and its name, along with a picture of the storefront. Down the street, she notices a series of new crosswalks at a complex intersection. Entering them on her mobile would be tedious, so she decides to take a picture that she will use later at home to update the map from her laptop (Figure 7).

In this scenario, text annotations (restaurant name, type and address) have been entered using a soft keyboard. But the smartglasses have also proven useful to capture other types of annotations in a streamlined and comfortable manner. Moreover, all of these annotations (tag, storefront picture, panoramic picture) can be displayed as if floating in the air around the smartphone. Here again, this saves screen space on the small handheld device, which this time displays a very dense and complex widget (the map). Offloading into the air also lets Laura quickly access and review all her annotations, since these can be represented as preview thumbnails.

*5.1.3 $S_{video}$ – Annotating a video (Figure 8).* Valentin is a CGI-movie enthusiast. He likes to share his videos with fellow animators to get feedback, and to give feedback about their own creations. One of his friends just sent him a video in which a spaceship approaches a star system. Valentin only has his tablet at hand along with his smartglasses.

Valentin plays the video on the tablet. He then scrubs through it using the interactive timeline in order to find the frames and time intervals in which he wants to make a series of suggestions. For instance, he selects a sequence in which the ship enters a planet’s orbit, and suggests making that clip faster. He then selects a still frame in which the star is visible and suggests adjusting its color.

In this scenario, the tablet’s pencil lets Valentin easily circle elements in the video such as the star and write the corresponding comments on a canvas to create post-it notes. The tablet shows

[6]https://www.openstreetmap.org

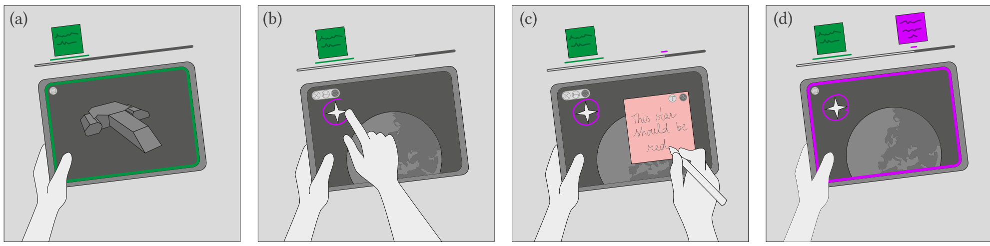



Fig. 8. $S_{video}$ scenario: a) Valentin has already entered one first handwritten Post-It-like annotation using his stylus for the time interval shown in green in the timeline; b-c) he adds another annotation later in the video, suggesting that the star's color be changed to red; d) the new annotation now appears in the timeline, with its temporal scope shown in purple.

the video only, while annotations are displayed floating in the air above the tablet. Displaying the annotations in the air does not occlude the video itself. It also makes it possible to arrange them temporally along a timeline that is also floating in the air, which better shows the distribution and scope (time interval) of annotations across the video.

## 6 DISCUSSION

As illustrated in the above scenarios, the forthcoming generation of smartglasses have the potential to improve the user experience when annotating various types of digital content on a mobile device. Not only can they improve the presentation of annotations by offloading them into the air using AR, but they also enable new ways to capture annotations thanks to their built-in camera. However, there are limitations to the approach that we discuss below. Some limitations, captured in our participants' feedback, are primarily related to capabilities of current AR technologies in terms of input and output. Another limitation – which is rather a privacy-related question – relates to capturing content from the external environment inconspicuously with AR eyewear.

### 6.1 Limitations to Interaction with Content in Mid-air

Some study participants asked if the annotations could be made more interactive than simply getting larger when hovering. For instance, they mentioned using the annotations as navigation helpers, to quickly jump to a specific location. Or they would have liked to use the timeline displayed in AR to be interactive. However, techniques for bare-hand input with current AR eyewear are too rudimentary. For instance, with the Microsoft Hololens 2 used to prototype *AnnotAR*, users can directly *tap* components that float in the air or rely on raycast pointing (based on gaze or hand input) and make selections through air-tap gestures. While this might be enough to select an annotation thumbnail, such in-air interaction is typically tiring and lacks precision [6, 22]. More elaborate interactions such as navigating the timeline or precisely adjusting the size of an annotation thumbnail would require implementing more advanced input techniques (in the spirit of the high-precision widgets proposed in [7]). Such size adjustments could be performed with bi-manual techniques as well, when the device rests on, *e.g.*, a surface.

Additionally, depending on the AR hardware considered, bare-hand input could be replaced or combined with other modalities. For instance, a physical trigger on the side of the smartglasses – as when pressing the camera button on Google Glass – could be a nice way of invoking discrete actions such as taking a picture (illustrated in Figure 1-b). Other input modalities, such as voice, can be considered as well when socially acceptable.

### 6.2 Limitations to the Visualization of Content in Mid-air

Feedback from our study also reveals that some participants envision a system that displays more visual elements in the air. For instance, some participants would like annotations whose scope is no longer (or not yet) in the current viewport to still be visible. Another participant suggested reinforcing the association between an annotation and its scope with graphical arrows explicitly connecting them. Such feedback implies that the annotation layout strategy needs to be iterated upon, if only to support the simultaneous display of a higher number of annotations in the air. In particular, future AR hardware will likely offer a wider field-of-view than that of the Hololens 2. However, how to efficiently make use of an extended field-of-view remains an open question. Studies such as Bahna & Jacob's about Augmented Reading [2] suggest that users can benefit from having information readily available in their peripheral vision, but the advantage of widening the field-of-view in the specific case of *AnnotAR* is not clear. A wide field-of-view can actually be beneficial to information-seeking tasks in an AR-only scene [39] but increasing the distance between content displayed on a handheld device and content displayed in the air could also degrade user performance for tasks that involve looking at both [3].

### 6.3 Privacy

Smartglasses let users take pictures and videos in a rather inconspicuous manner, which obviously raises privacy issues. This is a general problem with smartglasses [24, 25] that goes well beyond the specific case of annotating digital content. People cannot easily notice that they are being recorded [26], at least not as easily as when a smartphone is used to make the recording. The first two scenarios illustrate situations where this might be a problem: for the professor in scenario $S_{text}$, and for bystanders in the street in scenario $S_{map}$. Issues relate to both user behavior and malicious applications, as discussed in the rich literature on this research topic – see, *e.g.*, [13, 28] for a representative set of security and privacy concerns as well as [14, 23] for ways to address them from an HCI perspective.

## 7 SUMMARY AND FUTURE WORK

We prototyped AR enhancements in a proof-of-concept annotation system which we used as a probe to collect user feedback about the viability and utility of an AR-enhanced annotation experience. Participants gave generally-favorable feedback. Answers to the SUS questionnaires rated the usefulness and usability of the new features positively. We observed that it was easy for participants to relate mid-air annotations with the corresponding snippets in the primary content, despite the fact that the association is encoded using a fairly implicit visual channel (color hue).

We also prototyped a set of scenarios to illustrate the possibilities that an AR-enhanced annotation system opens. Mobile devices can of course already take pictures and videos. But a smartglasses' camera will enable capturing pictures and videos *from the user's perspective*, *leaving their hands free* for other actions *while recording* such as, *e.g.*, performing deictic pointing gestures or manipulating physical objects. This way of capturing pictures and videos without having to hold a device also lets users adopt a more relaxed posture. In addition, it can also be socially less awkward, depending on the context.

AR smartglasses are still in their infancy, however. Even if tremendous progress has been made, the hardware remains clumsy and the field of view is still limited. There is room for improvement on the software side as well: graphics rendering, interaction techniques and spatial registration, to name a few. All these factors impact the user experience and likely have had an adverse effect on participants' evaluation of the approach in our preliminary study.

Finally, participants also asked if the annotations could be made more interactive, which suggest interesting ideas to investigate as future work. Pushing the idea of interactive in-air annotations further, we can even think of making them *active*. For example, systems like XLibris [17], Ink-Seine [21] or the recent Holodoc [29] can turn annotations into actions to fetch and embed related external information in the content of interest. The ActiveInk system [40] also lets users turn annotations made on visualizations into commands that can act on the underlying data (such as, *e.g.*, filtering). In the same spirit, in-air annotations could trigger different actions. For example, an annotation that was captured with smartglasses could be used to generate queries in order to gather information about the location where this capture was made. The possibilities are quite broad, but require AR hardware and core annotation system features to mature before they can be deeply explored and reach their full potential.

## ACKNOWLEDGMENTS

This work was supported by French government funding managed by the National Research Agency under the Investments for the Future program (PIA) grant ANR-21- ESRE-0030 (CONTINUUM)

## REFERENCES

[1] Maneesh Agrawala and Michael Shilman. 2005. DIZI: A Digital Ink Zooming Interface for Document Annotation. In *Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction (INTERACT'05)*. Springer-Verlag, 69–79. https://doi.org/10.1007/11555261_9

[2] Eric Bahna and Robert J. K. Jacob. 2005. Augmented Reading: Presenting Additional Information without Penalty. In *CHI '05 Extended Abstracts*. ACM, 1909–1912. https://doi.org/10.1145/1056808.1057054

[3] Sunyoung Bang, Hyunjin Lee, and Woontack Woo. 2020. Effects of Augmented Content's Placement and Size on User's Search Experience in Extended Displays. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 184–188. https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00056

[4] Patrick Baudisch and Ruth Rosenholtz. 2003. Halo: A Technique for Visualizing off-Screen Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, 481–488. https://doi.org/10.1145/642611.642695

[5] Blaine Bell, Tobias Höllerer, and Steven Feiner. 2002. An Annotated Situation-Awareness Aid for Augmented Reality. In *Proceedings of the Symposium on User Interface Software and Technology (UIST '02)*. ACM, 213–216. https://doi.org/10.1145/571985.572017

[6] Eugenie Brasier, Olivier Chapuis, Nicolas Ferey, Jeanne Vezien, and Caroline Appert. 2020. ARPads: Mid-air Indirect Input for Augmented Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 332–343. https://doi.org/10.1109/ISMAR50242.2020.00060

[7] Eugenie Brasier, Emmanuel Pietriga, and Caroline Appert. 2021. AR-enhanced Widgets for Smartphone-centric Interaction. In *Proceedings of the Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '21)*. ACM, 19 pages. https://doi.org/10.1145/1122445.1122456

[8] Nicki Brower. 2021. Apple - Adopt Quick Note. https://developer.apple.com/videos/play/wwdc2021/10264/.

[9] A. J. Bernheim Brush, David Bargeron, Anoop Gupta, and J. J. Cadiz. 2001. Robust Annotation Positioning in Digital Documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, 285–292. https://doi.org/10.1145/365024.365117

[10] Rahul Budhiraja, Gun. A. Lee, and Mark Billinghurst. 2013. Using a HHD with a HMD for mobile AR interaction. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1–6. https://doi.org/10.1109/ISMAR.2013.6671837

[11] Yun Suk Chang, Benjamin Nuernberger, Bo Luan, and Tobias Höllerer. 2017. Evaluating gesture-based augmented reality annotation. In *IEEE Symposium on 3D User Interfaces (3DUI)*. 182–185. https://doi.org/10.1109/3DUI.2017.7893337

[12] M. A. Chatti, T. Sodhi, M. Specht, R. Klamma, and R. Klemke. 2006. u-Annotate: An Application for User-Driven Freeform Digital Ink Annotation of E-Learning Content. In *IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. 1039–1043. https://doi.org/10.1109/ICALT.2006.1652624

[13] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. 2019. Security and Privacy Approaches in Mixed Reality: A Literature Survey. *ACM Comput. Surv.* 52, 6, Article 110 (oct 2019), 37 pages. https://doi.org/10.1145/3359626

[14] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-Mediating Technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2377–2386. https://doi.org/10.1145/2556288.2557352

[15] Anna Eiberger, Per Ola Kristensson, Susanne Mayr, Matthias Kranz, and Jens Grubert. 2019. Effects of Depth Layer Switching between an Optical See-Through Head-Mounted Display and a Body-Proximate Display. In *Symposium on Spatial User Interaction (SUI '19)*. ACM, Article 15, 9 pages. https://doi.org/10.1145/3357251.3357588

[16] Gene Golovchinsky and Laurent Denoue. 2002. Moving Markup: Repositioning Freeform Annotations. In *Proceedings of the Symposium on User Interface Software and Technology (UIST '02)*. ACM, 21–30. https://doi.org/10.1145/571985.571989

[17] Gene Golovchinsky, Morgan N. Price, and Bill N. Schilit. 1999. From Reading to Retrieval: Freeform Ink Annotations As Queries. In *Proceedings of the SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*. ACM, 19–25. https://doi.org/10.1145/312624.312637

[18] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. MultiFi: Multi Fidelity Interaction with Displays On and Around the Body. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 3933–3942. https://doi.org/10.1145/2702123.2702331

[19] Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. 2008. Wedge: Clutter-Free Visualization of off-Screen Locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, 787–796. https://doi.org/10.1145/1357054.1357179

[20] Mark Harrower and Cynthia Brewer. 2003. ColorBrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.

[21] Ken Hinckley, Ken Hinckley, Shengdong Zhao, Raman Sarin, Patrick Baudisch, Edward Cutrell, Michael Shilman, and Desney Tan. 2007. InkSeine: In Situ Search for Active Note Taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, 251–260. https://doi.org/10.1145/1240624.1240666

[22] Brett Jones, Rajinder Sodhi, David Forsyth, Brian Bailey, and Giuliano Maciocci. 2012. Around Device Interaction for Multiscale Navigation. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) *(MobileHCI '12)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2371574.2371589

[23] Marion Koelle, Swamy Ananthanarayan, Simon Czupalla, Wilko Heuten, and Susanne Boll. 2018. Your Smart Glasses' Camera Bothers Me! Exploring Opt-in and Opt-out Gestures for Privacy Mediation. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (Oslo, Norway) *(NordiCHI '18)*. Association for Computing Machinery, New York, NY, USA, 473–481. https://doi.org/10.1145/3240167.3240174

[24] Marion Koelle, Abdallah El Ali, Vanessa Cobus, Wilko Heuten, and Susanne CJ Boll. 2017. *All about Acceptability? Identifying Factors for the Adoption of Data Glasses.* ACM, 295–300. https://doi.org/10.1145/3025453.3025749

[25] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't Look at Me That Way! Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, 362–372. https://doi.org/10.1145/2785830.2785842

[26] Marion Koelle, Katrin Wolf, and Susanne Boll. 2018. Beyond LED Status Lights - Design Requirements of Privacy Notices for Body-Worn Cameras. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (Stockholm, Sweden) *(TEI '18)*. Association for Computing Machinery, New York, NY, USA, 177–187. https://doi.org/10.1145/3173225.3173234

[27] Ricardo Langner, Marc Satkowski, Wolfgang Büschel, and Raimund Dachselt. 2021. MARVIS: Combining Mobile Devices and Augmented Reality for Visual Data Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Article 468, 17 pages. https://doi.org/10.1145/3411764.3445593

[28] Sarah M. Lehman, Abrar S. Alrumayh, Kunal Kolhe, Haibin Ling, and Chiu C. Tan. 2022. Hidden in Plain Sight: Exploring Privacy Risks of Mobile Augmented Reality Applications. *ACM Trans. Priv. Secur.* 25, 4, Article 26 (jul 2022), 35 pages. https://doi.org/10.1145/3524020

[29] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. 2019. HoloDoc: Enabling Mixed Reality Workspaces That Harness Physical and Digital Content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1–14. https://doi.org/10.1145/3290605.3300917

[30] Catherine C. Marshall. 1997. Annotation: From Paper Books to the Digital Library. In *Proceedings of the Second ACM International Conference on Digital Libraries (DL '97)*. ACM, 131–140. https://doi.org/10.1145/263690.263806

[31] Fabrice Matulic and Moira C. Norrie. 2012. Supporting Active Reading on Pen and Touch-operated Tabletops. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '12)*. ACM, 612–619. https://doi.org/10.1145/2254556.2254669

[32] Alexandre Millette and Michael J. McGuffin. 2016. DualCAD: Integrating Augmented Reality with a Desktop GUI and Smartphone Interaction. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 21–26. https://doi.org/10.1109/ISMAR-Adjunct.2016.0030

[33] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. 2019. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, 1–11. https://doi.org/10.1145/3290605.3300815

[34] M. R. Morris, A. J. B. Brush, and B. R. Meyers. 2007. Reading Revisited: Evaluating the Usability of Digital Display Surfaces for Active Reading Tasks. In *IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07)*. 79–86. https://doi.org/10.1109/TABLETOP.2007.12

[35] Erwan Normand and Michael J. McGuffin. 2018. Enlarging a Smartphone with AR to Create a Handheld VESAD (Virtually Extended Screen-Aligned Display). In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 123–133. https://doi.org/10.1109/ISMAR.2018.00043

[36] Kenton O'Hara and Abigail Sellen. 1997. A Comparison of Reading Paper and On-line Documents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, 335–342. https://doi.org/10.1145/258549.258787

[37] Beryl Plimmer, Samuel Hsiao-Heng Chang, Meghavi Doshi, Laura Laycock, and Nilanthi Seneviratne. 2010. iAnnotate: Exploring Multi-user Ink Annotation in Web Browsers. In *Proceedings of the Australasian Conference on User Interface (AUIC '10)*. 52–60. https://doi.org/10.5555/1862280.1862289

[38] Jing Qian, Qi Sun, Curtis Wigington, Han L. Han, Tong Sun, Jennifer Healey, James Tompkin, and Jeff Huang. 2022. Dually Noted: Layout-Aware Annotations with Smartphone Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 552, 15 pages. https://doi.org/10.1145/3491102.3502026

[39] Donghao Ren, Tibor Goldschwendt, YunSuk Chang, and Tobias Höllerer. 2016. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *IEEE Virtual Reality*. 93–102. https://doi.org/10.1109/VR.2016.7504692

[40] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. 2019. ActiveInk: (Th)Inking with Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1–13. https://doi.org/10.1145/3290605.3300272

[41] Hugo Romat, Emmanuel Pietriga, Nathalie Henry-Riche, Ken Hinckley, and Caroline Appert. 2019. SpaceInk: Making Space for In-Context Annotations. In *Proceedings of the Symposium on User Interface Software and Technology (UIST '19)*. ACM, 871–882. https://doi.org/10.1145/3332165.3347934

[42] Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., 249–256. https://doi.org/10.1145/274644.274680

[43] Marcos Serrano, Dale Hildebrandt, Sriram Subramanian, and Pourang Irani. 2014. Identifying Suitable Projection Parameters and Display Configurations for Mobile True-3D Displays. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14)*. ACM, 135–143. https://doi.org/10.1145/2628363.2628375

[44] Craig S. Tashman and W. Keith Edwards. 2011. LiquidText: A Flexible, Multitouch Environment to Support Active Reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, 3285–3294. https://doi.org/10.1145/1978942.1979430

[45] Dirk Wenig, Johannes Schöning, Alex Olwal, Mathias Oben, and Rainer Malaka. 2017. WatchThru: Expanding Smartwatch Displays with Mid-Air Visuals and Wrist-Worn Augmented Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, 716–721. https://doi.org/10.1145/3025453.3025852

[46] Jason Wither, Stephen DiVerdi, and Tobias Höllerer. 2009. Annotation in outdoor augmented reality. *Computers & Graphics* 33, 6 (2009), 679–689. https://doi.org/10.1016/j.cag.2009.06.001

[47] Dongwook Yoon, Nicholas Chen, and François Guimbretière. 2013. TextTearing: Opening White Space for Digital Ink Annotation. In *Proceedings of the Symposium on User Interface Software and Technology (UIST '13)*. ACM, 107–112. https://doi.org/10.1145/2501988.2502036

[48] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: Blending Ink, Speech, and Gesture to Support Collaborative Document Review. In *Proceedings of the Symposium on User Interface Software and Technology (UIST '14)*. ACM, 481–490. https://doi.org/10.1145/2642918.2647390

[49] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 1–14. https://doi.org/10.1145/3313831.3376233