Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

ARtention: A Design Space for Gaze-adaptive User Interfaces in Augmented Reality

Ken Pfeuffer[a], Yasmeen Abdrabou[a], Augusto Esteves[b,*], Radiah Rivu[a], Yomna Abdelrahman[a], Stefanie Meitner[c], Amr Saadi[d], Florian Alt[a]

[a]*Bundeswehr University Munich, Munich, Germany*
[b]*ITI / LARSyS, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal*
[c]*LMU Munich, Munich, Germany*
[d]*German University in Cairo, Cairo, Egypt*

ARTICLE INFO

Article history:
Received January 10, 2021

Keywords: Augmented reality, AR, Mixed reality, MR, Gaze Interaction, Attention, Visualization, Design space

ABSTRACT

Augmented Reality (AR) headsets extended with eye-tracking, a promising input technology for its natural and implicit nature, open a wide range of new interaction capabilities for everyday use. In this paper we present ARtention, a design space for gaze interaction specifically tailored for in-situ AR information interfaces. It highlights three important dimensions to consider in the UI design of such gaze-enabled applications: transitions from reality to the virtual interface, from single- to multi-layer content, and from information consumption to selection tasks. Such transitional aspects bring previously isolated gaze interaction concepts together to form a unified AR space, enabling more advanced application control seamlessly mediated by gaze. We describe these factors in detail. To illustrate how the design space can be used, we present three prototype applications and report informal user feedback obtained from different scenarios: a conversational UI, viewing a 3D visualization, and browsing items for shopping. We conclude with design considerations derived from our development and evaluation of the prototypes. We expect these to be valuable for researchers and designers investigating the use of gaze input in AR systems and applications.

1. Introduction

Augmented reality (AR) is a maturing technology poised to become a popular personal device for everyday use [1]. AR enables the projection of context-relevant information anchored to the real world into our view, facilitating spatial tasks and in-situ knowledge generation and understanding [2]. Recently, Lu et al. introduced the *Glanceable AR* paradigm aimed at user interfaces that can provide information at any time and place, with interaction as effortless as a glance [3]. We investigate this paradigm for temporary information interfaces in the world, aimed to reside in the background of our attention so we can fully attend reality, but transition to the fore when relevant to our interests [4, 5]. Within the reality-virtuality (RV) continuum, this involves frequent moving between the real environment and virtual interfaces in AR [6].

One modality with promising potential to enable such transitions, directly related to attentional processes, is our eye gaze. Our eyes are always available, are easy to track, and implicitly tell what we are interested in; which fits to properties suitable for everyday AR usage. Research in this direction so far has employed gaze as a technique to reveal AR information [7, 8, 9], and to trigger object selection by dwell-time (with respect to gaze-only interaction techniques) [10, 11]. However, dwell-time has been originally developed for high-performance screen selections, a very different nature to the open and potentially information-rich AR environments. To complement prior

*Corresponding author: Tel.: +351-218-419-910
e-mail: augusto.esteves@tecnico.ulisboa.pt (Augusto Esteves)

art, we investigate combining interaction tasks of revealing, traversing and/or selection of content. This yields a novel challenge, that is how to optimize dwell-time and gaze interaction for a wider range of user intents in the context of AR.

In this paper we present ARtention, a design space for gaze-input based interfaces for AR applications. This space covers three dimensions critical for designers and researchers in this space (see Figure 1). First, understanding the role of gaze as a proxy for our engagement with real-world or AR content – *reality-virtuality (RV) continuum transitions* – how can we quickly access AR information, and how can we seamlessly return to the task-at-hand? Second, understanding the role of gaze as a proxy for attention on specific AR information – *information level transitions* – how can AR systems hint at available information, and continuously unfold this in response to our attention? And third, understanding how mechanisms such as dwell-time can still coexist in this space to offer both information level transitions and effective user interface (UI) selections that explicitly alter the system state – *task transitions*.

To illustrate the three dimensions of this design space, we have implemented and studied three prototype applications that highlight different user needs in the context of AR. Each application exemplifies how particular transitions can be supported, and their evaluations allow initial insights regarding general usability to be collected as well as suitable gaze-adaptive timings for revealing and selecting AR information to be identified:

- **Conversational UI:** The first application illustrates a classical conversational interface where the person we are talking to is augmented with relevant information (e.g., name, interests). In this context our focus is on the person we are talking to, and transitions in and out of AR need to take place quickly and effortlessly. Our prototype also explores transitions between three AR information levels, in a study where participants (N=12) have this conversational experience with two confederates.

- **Tree of Life:** The second application illustrates deeper information level transitions by proving us with a large 3D visualization of evolutionary biology that carefully unfolds in response to our attention to particular nodes. This was studied again with 12 participants, and assessed from an user experience point-of-view.

- **Shopping:** The final probe presents a supermarket-like shopping interface that relies on gaze to provide AR information on the items on display (e.g., cost, origin), but also item selections for purchasing. The study of this probe involved various dwell-time implementations, and we report task completion times and error rates.

Contribution Statement: Our contribution is threefold: (1) a design space that articulates the role of gaze in AR across three factors – RV continuum, information level, and task transitions; (2) the design and implementation of three gaze-based AR prototype applications; and (3) the study and discussion of these applications, illustrating how can designers and researchers think of AR in the context of our design space.

2. Related Work

Our work draws directly from prior art on adaptive AR, gaze interaction, and glanceable AR.

2.1. Adaptive AR

Designing for superimposed information on real environments involves information visualization and placement and has long been a significant challenge in AR [12]. Environments can have many different buildings, streets, and other objects, in addition to various UI elements – each potentially offering numerous information. To avoid experiencing information overload, systems can be designed to contain explicit and implicit means to organize interfaces, each having various pros and cons [13].

Explicit input, e.g. through a pointing gesture [14], will provide us with the exact information we desire. However, in everyday settings it can become tedious to interact with many objects, and the need for an input device or hand gesture may not be desired. On the other hand, an implicit method avoids manual effort by designing for a more proactive behavior where information is revealed based on proxemics [15] or context. Context factors can include environmental knowledge (object location/shape), or goals and rules to approximate relevance [16, 17]. Various AR systems were developed to follow this approach [18, 19, 20]. A limitation of the implicit method is the approximated relevance, which need not necessarily be correlated with our real intent. An ideal interface is designed to support both means to show approximated relevant information yet provide manual means to interact with it, which we argue is ideal to accomplish through gaze in consideration of various design factors.

Others have addressed this problem by presenting a series of strategies on how to adapt AR information based on our attention and their surroundings [21]. While insightful, this work focuses on the technical challenges of spatially rearranging AR content in a 3D scene, models attention via user orientation, and was studied and motivated by specific mobility scenarios. These and other technical implementations (e.g., [22]) are critical to the field, but our contribution rests on a deeper understanding of user attention via gaze as a rich mechanism to not only adapt AR content, but facilitate user transitions to and from AR in everyday tasks that require interaction with both physical and digital elements.

2.2. Gaze Interaction

In the HCI literature, the gaze modality is often considered as an explicit input modality. Bolt's Gaze-orchestrated dynamic windows demonstrated how gaze can be leveraged for a better organization of visual content [23], i.e. "*to help the observer cope with the onslaught of events on the one hand, yet enable on the other hand continued close contact with that ever changing ensemble*". Gaze therefore has the potential to provide a dual advantage for information-rich interfaces (as AR): it can avoid information overload and provide information on demand.

Gaze as interaction technique has been explored in Jacob's seminal paper where we select UI elements by looking at them [24]. One of the main challenges identified was the 'Midas

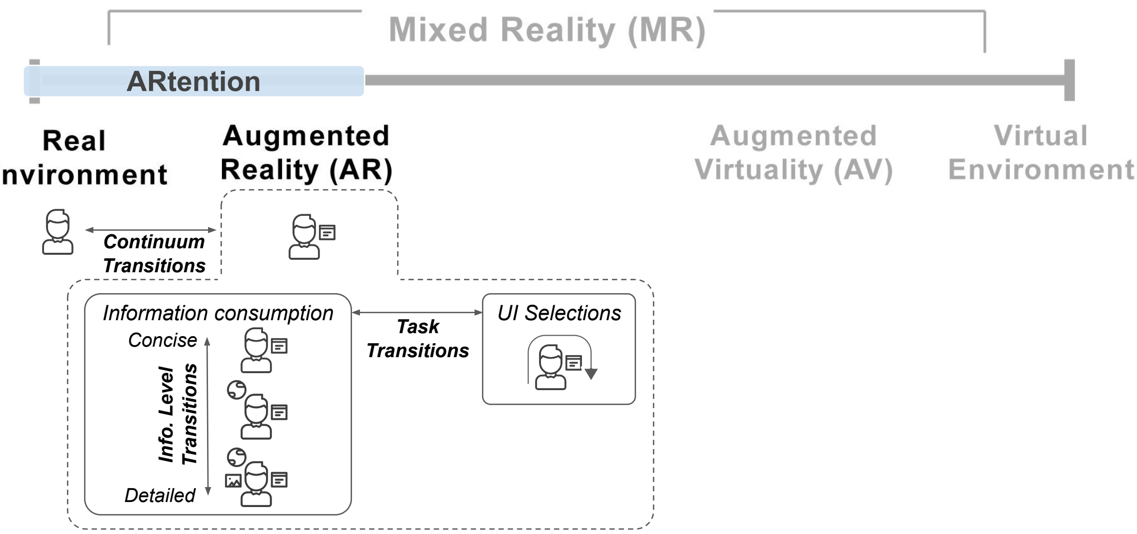

Fig. 1. The ARtention design space in the context of the Reality-Virtuality (RV) continuum, illustrating the role of gaze across three dimensions: attending to reality or to AR content (*continuum transitions*), consuming such content or performing UI selections (*task transitions*), or continuously unfolding more content in response to user's attention and interest (*information level transitions*).

touch', i.e. the ambiguity to distinguish simple 'looking' from 'selection' via gaze. The dwell-time confirmation technique attempts to addresses this, where selection is confirmed after a specific viewing time. Similar approaches can be found in a wide range of systems, including Vertegaal's Attentive User Interfaces [5] or Quarfordt and Zhai's use of gaze intensity to infer interest during interaction with a tourist information application [25]. In cases where dwell-time is not suitable (e.g., too long wait time), a additional manual input can confirm gaze selection, such as keyboard and mouse [24, 26, 27], touchscreen [28, 29], or gesture [30]. This approach has been extended to object selection in 3D environments by introducing gaze and controller techniques [31, 11], a combination of gaze and freehand gestures for object manipulation [32] and gaze-enhanced menu interfaces [33]. In this work, we focus on gaze-only as a potentially implicit and effortless method. As the technology moves to the mainstream, we believe it is crucial to gain a holistic understanding of design factors in everyday AR. Along this line, Hirzle et al. recently introduced a design space for gaze interaction on head mounted devices, which includes AR [34]. Their work focuses on human depth perception and technical issues, which we extend with a perspective on gaze interaction tailored for everyday AR.

2.3. Gaze Interaction in AR

Research in immersive virtual environments exploited gaze as context feature to adapt content in the user interface. Ajanki et al. included features from objects, the environment, other people, and gaze as interest metrics to make decisions on how to organize virtual content in our field-of-view [7]. Esteves et al. explored the interaction with AR interface for smart homes and devices via gaze-based controls [35]. Park et al. used an advanced dwell-time approach based on gaze frequencies to trigger AR content [36]. Kim et al. divided an AR object in three areas that are each selectable with dwell-time [10]. McNamara et al. built a prototype where object labels are placed according to gaze information [9], as well as a VR system where information is revealed when looking at the region around the target [37, 38]. Finally, Esteves et al. provided a in-depth comparison of selections techniques for AR and VR, including various gaze-based approaches [39]. These works show that use of gaze to activate information is promising to provide a novel way to interact with lightweight AR content. We take this activation as a first step, and extend this with more advanced interaction tasks through our design space.

Our work relates to Glanceable AR interfaces introduced by Lu and Davari et al. [3, 8]. We share the focus on temporary accessible information, and how our visual attention can reveal/conceal temporary information. However, we consider world-fixed interfaces, which yields different challenges than a personal UI. The main distinction is our focus on transitions. We investigate the UI factors of primary/secondary focus, and single-task/multi-task factors not separate, but potentially in unison by designing for transitions. Finally, we expand our understanding of the occlusion techniques described by Davari et al. [8] to address the spatial positioning and translucency of AR information. These were studied in a task resembling our first prototype application, where our task resides primarily in the real world and AR content is to be consumed sparingly and at a glance. We designed two other applications where our primary task gradually takes place further from reality to explore this interplay between occlusion and information access.

3. The ARtention Design Space

The ARtention design space involves three dimensions: RV Continuum, Information Level, and Task transitions (see Figure 1). In the following we detail these dimensions, and how gaze can be leveraged for not only interaction, but as a mechanism to transition between these dimensions.

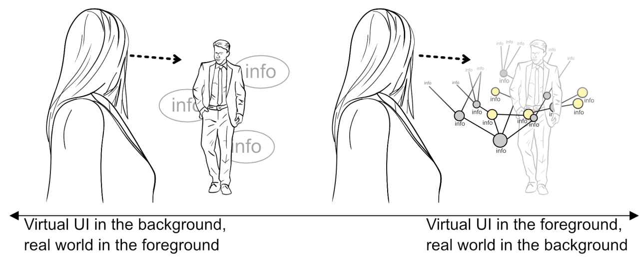

Fig. 2. Depending on the application context, the UI considers the ratio between reality and virtuality in response to background and foreground attention. For example, in a real-world conversation the virtual annotations will reside primarily in the background of the field of view; while in an educational visualization, the interface will reside primarily at the fore.

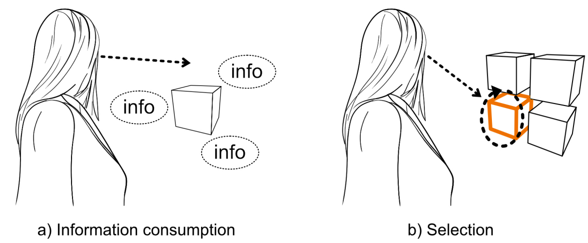

Fig. 3. Gaze interaction can be considered from two perspectives. *Consumption* is about revealing and seeing lightweight information – the system analyses gaze in the background, and can adapt the content. During *selections* users decide to voluntarily select a target normally via dwell.

3.1. RV Continuum Transitions

The RV continuum spans all possible compositions between real and virtual objects [6]. Our attention describes whether we are currently focusing on a real-world task, or on virtual content. Prior work focused on techniques that prioritize one or the other [8], but the advantage of gaze interaction is that we can easily transition between prioritizing what is real and what is virtual. There can be applications where transitions occur more frequently, and others where it is less common. Based on the expected frequency of these transitions, and the ratio of time spent in real or virtual, we can inform the design of a gaze-based AR interface regarding how much it should remain in the foreground or background of our attention. We describe two contrasting scenarios as examples for this dimension.

3.1.1. Reality-prioritized

When the virtual UI should remain mostly in the background of our attention, it should be designed subtly and with minimal potential distraction. For example, we could be talking to another person and occasionally access some information about a particular topic of the conversation (Figure 2-left). Initially, the UI should show indications that are so subtle that they do not really change the user's perspective of its environment. However, if looking precisely at them, they are easy to grasp and can even become interactive. When the user's gaze lands on such an indication, the corresponding content fully reveals itself and is easy to interpret. In addition, if our gaze has left the interface, the state of the interface can return to the previous indicative state to minimize the potential for distraction.

3.1.2. Virtuality-prioritized

On the other hand, other types of content do not necessarily have to fade away when we avert our gaze; on the contrary, it might be important to keep history of what has been viewed so we are aware of our process, and also can visually compare different information. Figure 2-right illustrates an example where such a visualization is overlaying the real environment if necessary for the virtual task. In other scenarios this AR content can respond to not only gaze, but to the relation between the user's body and the world or task such as looking at a building or item from a different perspective, or the manipulation of a physical artifact (see Figure 4).

3.2. Information Level Transitions

This dimension refers to the information content available in levels that users can traverse via gaze interaction. The dimension extends the concise vs. detailed factor of a single information layer, to the possibility to browse multiple layers to provide users a way to gradually consume more information. The UI design can depend on various trade-offs. This dimension is similarly considered in [3]. Such a feature is, for instance, useful for more open-ended and exploratory applications. In them, information can gradually appear to match our increasing interest and curiosity. We distinguish between a default single-layer, and the multi-layer information approaches.

3.2.1. Single-Layer Information

This refers to content we can quickly glance at to get informed about an element. Often it is the first point of a continuum transition to virtual interfaces. In the case we are not interested, the interface needs to implement methods that avoid clutter and are easy to ignore (e.g., a simple caption, iconic information). The first layer can also be an indicator, i.e., a visual element demonstrating us that there is a virtual interface available – but again should not be distracting if we decide to ignore it to focus on the real-world.

3.2.2. Multi-Layer Information

This refers to information where we can gain a deeper, more detailed understanding of a topic or element, for which we also spend more time and effort engaged with. It can start with a single-layer and continuously unfold onto further layers. Interfaces can be designed to offer two ways to access multi-layered information. First, directly providing a lot of information at once, for example, if it is important to access an overview or detailed information straight away. Second, via concise content that can be provided over time so as to not overload or clutter the real-world view, in parallel to allowing us to adapt to a continuous progression of information. Such progression can be adapted in response to our increased visual attention on the unfolding element based on our dwell times.

3.2.3. Transition

To transition from initial information to gradually more detail, the interface should take our visual attention into account in the process of presenting multiple layers of information. This

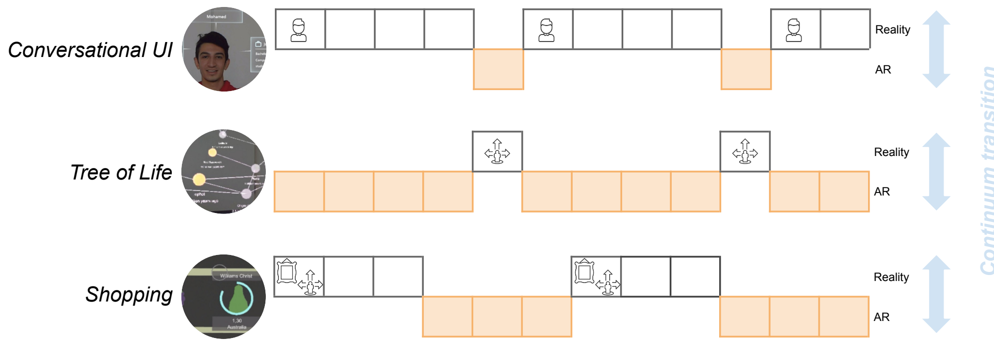

Fig. 4. Our three applications in the context of RV continuum transitions, highlighting the interplay of interactions taking place in reality and in AR. The Conversational UI illustrates an application where the main task takes places in the real-world, and we rely on AR information sporadically and for short amounts of time. The Tree of Life illustrates the opposite scenario, where we primarily attend to AR content and sporadically take action in the real-world (e.g., re-positioning us in space for a different perspective on this content). Finally, the Shopping application illustrates a more leveled approach, where we can spend as much time attending to the real-world (e.g., navigating the supermarket) as to AR information (e.g., browsing item information).

can be designed by temporal and spatial multiplexing of the UI items in relation to gaze data. A temporal-only approach adapts the content of an element after an estimated time of consumption of the information provided to us. Thus, the new information resides in the same spatial element, and can replace prior information. To avoid replacement, the information can be provided spatially offset. To establish a sense of connection between information layers, the position of the new content window can be in the vicinity of the previous one; allowing us to immediately understand and glance over to the next item.

3.3. Task Transitions

Gaze interfaces can be designed in consideration of two fundamental tasks in AR interaction: information consumption and UI selections. Both tasks are illustrated in Figure 3. We describe each, and the transitions between them.

3.3.1. Information Consumption

Here we call "Information consumption" as all the cases where one reads or engages with a picture in the contextual information displayed over an object in the AR interface. Many use cases have been studied where such a task is common, e.g., in a library environment [40, 41, 42], at the computer desk [18], or at a university campus [43]. A difference to a more formal selection task, is that dwelling on an object is not consequential in that it triggers an action. Rather, when the user looks at the UI, it can expand to take more screen estate to convey more information, but return to original state when the user stops looking.

Such a gaze-based revealing mechanism can be implemented using various timings to improve the look & feel of the task. It includes: (1) the wait times required to acknowledge our attention and reduce information overload and the flickering of elements when the eyes scan the UI; (2) the fade-in and fade-out times that support a better transition between showing/hiding elements and error recovery; and (3) stay time – the time an element will remain visible after we looks away, supporting information retrieval and peripheral awareness.

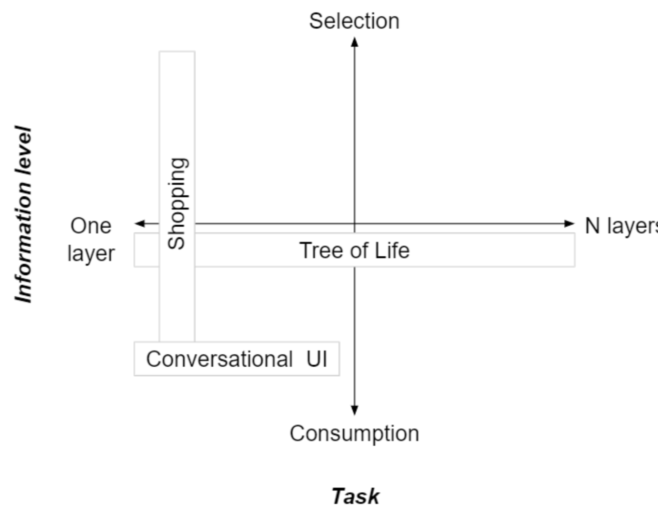

Fig. 5. A diagram showing how our applications map to the two design dimensions (information level and task). All of the applications have slightly different information consumption properties, but a particular unique case is the Shopping example, as it supports both tasks in the same interface for which we investigate the effect of different dwell timings. At the Information layer level, all applications starts with a basic layer; with the Conversational UI investigating three layers that respond to user glances. The Tree of Life demonstrates many layers and objects that gradually appear based on what one has seen before, and provides an interesting take on a storytelling-based visualization.

3.3.2. Selection

The most prominent selection method via gaze is the dwell-time mechanism, i.e. gazing at an object over a specific amount of time [24]. This is a different but important task to command the UI. As such, tasks are more consequential, changing their internal states. Duration of dwell depends on a speed/error trade-off. Timings for target selection can be considered from prior art, such as from Park et al. [44]. The literature on dwell times includes thresholds from 150 ms to 1500 ms [24, 45, 46, 47, 48, 49].

3.3.3. Transition

Both tasks can be intertwined when it is possible to estimate the time to have looked "enough" at some information to trigger a transition to another. This includes time to read and understand information on an element, and time to dwell-select an item without triggering accidental responses. The UI can rely on spatial mechanisms by positioning UI elements separate to selection buttons. This can depend on the functionality needed to cover the content. Simple descriptions of an object might not need selection functionality, whereas a more advanced interface can offer additional selection elements. For example, to share the position and text shown to one of your contacts; or even fully-fledged applications to provide users with detailed menus and visualizations that one can interact with via explicit gaze input.

4. Prototype Applications

We built three applications, illustrating how the ARtention design space can be leveraged to conceptualize and design a wide range of gaze-based AR experiences. Figure 4 illustrates how the three applications can be thought of in the context of RV continuum transitions; while Figure 5 provides an overview of how these applications map to the other two design dimensions (information level and task transitions). We describe the design of these applications in the context of ARtention, and informal evaluations that focus on the user experience and design parameters relevant to our design space. A screenshot of each application is provided: Conversational UI (Figure 6), Tree of Life (Figure 8), and Shopping (Figure 9). All applications were developed using a 1st generation Microsoft HoloLens (1268 × 720 px per eye, 60 Hz, 30° horizontal and 17.5° vertical FoV) with a binocular attachment from Pupil Labs to record eye data (200 Hz, 9-point calibration, 4.5 ms latency). The software was written in C# using the Unity Game Engine.

4.1. Application #1: Conversational UI

Continuum Transitions: this application illustrates an example of an AR system for tasks taking place primarily in the real-world (i.e., a conversation), where users rely on AR information sporadically and for short amounts of time (Figure 6) [50].

Information Level and Task Transitions: the application affords no time consuming UI selections (and thus no task transitions), and has solely *three information levels*. In the first, no AR content is present when the user is not attending to the conversation partner (0% opacity). In the second, a small outline of the AR information is shown when the user's gaze is on the person she/he is talking to (5% opacity), allowing them to understand the location of the relevant AR content, while remaining minimally invasive if they choose to ignore it. In the third, AR information is provided when gazed-upon (100% opacity) – this includes the person' name (top), current projects they are working on (left), and their job description (right). The transition between these information levels is performed via smooth fade in (0.5s) and fade out (0.8s) animations.

4.1.1. Evaluation

We evaluate users' experience of revealing and consuming virtual information during a real-world conversation. We scripted a conversation of three people of which one was the actual participant (wearing the AR HMD) and the two others confederates. Since users sat in fixed locations, we displayed the augmented information at absolute positions in the vicinity of the confederates' heads. The task was to imagine a first day at university and getting to know new colleagues. The study ended with a usability questionnaire and interviews where we asked participants to reflect on the conversational experience.

12 participants (6F) aged between 21 and 30 years (M = 25, SD = 2.99) took part in the study, mostly STEM students. Using a 3-point Likert scale (higher is better), users were asked to rate their experience with VR (M = 2.2, SD = 0.7), mobile AR (M = 2.2, SD = 0.7), head-mounted AR (M = 1.7, SD = 0.9), and eye-tracking (M = 2.2, SD = 0.8). Three participants wore contact lenses and two conducted the study without their glasses, with slightly impaired vision. Figure 7 depicts the study setup.

4.1.2. Results

User Experience. Our observations, questionnaire and interview reveal that participants welcomed the idea of having gaze-revealing information during real-world conversations. We categorize our insights as follows: participants described the UI as useful as it allows them to better remember information, especially when one has to remember names of people (P4, P8, P9) and "it was nice to see information about someone before they themselves say it" (P8). To one participant this design seemed to be useful only in specific applications such as when training or teaching, rather than during everyday conversations (P5). Few participants felt that from a social perspective, it can be overwhelming to see information in advance, especially given they are unfamiliar with these types of interfaces. Participants who shared similarities such as hobbies or same course of study found the conversation became easier and more natural (N=7).

Scenarios. Most participants suggested this type of system to be helpful for meeting new people (P2, P3, P4, P9) and well suited for a job interview (P3, P6, P9, P11), as they could gather information about the background of the interviewer and thus be "more confident and open in the interview" (P3). Additional information would also be conceivable at events such as a conference or meeting where many people come together, and useful for people "with poor memory or even with dementia" (P7).

Transitions. Overall we found that users preferred the gaze based transition concept due to this being less distracting (P1, P3, P7, P8, P11), in comparison to showing all AR information constantly (one of the study conditions). The reasons for this include that information overload can occur during these types of conversational scenarios, and a more selective approach using gaze can mitigate the issue. Yet, several participants preferred to see all AR information at all times. This was in part due to technical limitations with our system which resulted in inaccurate gaze tracking and erroneous system behavior. This additional gaze pointing effort made it "harder to concentrate on the conversation" (P1). One user stated that "having to select a text box by looking at it required more attention" (P10).

4.2. Application #2: Tree of Life

The Tree of Life is a large 3D visualization that supports in-situ exploration of complex 3D visualizations which can benefit from AR technology [51, 52, 53]. It is based on the concept

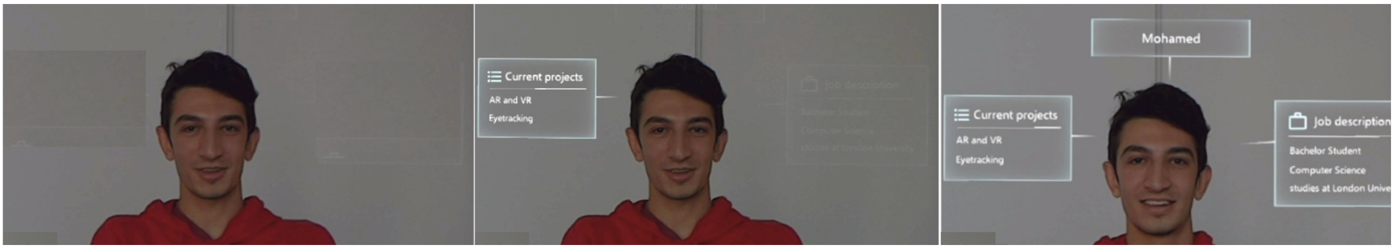

Fig. 6. Conversational UI: gaze provides a mean to reveal information on demand, allowing us to focus on the person we are talking to (left), while quickly revealing relevant AR information at a glance (middle). Having this content always-on could end up being distracting to us (right).

Fig. 7. The study setup for the Conversational UI application, where two confederates talk to the participant wearing the AR device.

of a Tree of Life visualization that describes how life evolved across species – a popular type of visualization seen in museums and art galleries [54, 55, 56]. The application includes one root, 21 internal nodes, and 54 leaf nodes. The whole model encompassed an area 2.5 m wide by 3.6 m high, and was displayed approx. 0.8 m below participants' heads – allowing them to walk around and comfortably explore any element of the tree. These elements have a radius of 7 cm, and distances between elements range from about 30 cm to 1 m. New elements unfolded upwards, at 10 cm increments. Nodes were presented as colored spheres and the leaves as 3D models of an animal species. Orange and gray represented folded and unfolded nodes, respectively. Visual feedback is provided by smoothly changing the sphere size when gaze hits one of the internal nodes, whereby the feedback for the leaf objects is a slow rotation animation. Initially, all elements but the root are folded. Figure 8 shows an example of the application in use, where a partially unfolded tree is presented. The user is free to walk around the visualization that is fixed to the middle of a room.

Continuum Transitions: the second prototype application presents a contrasting example to the previous prototype, where the user's task now revolves primarily around digital content. The only interaction taking place in relation to the real environment is users' movement in space, crouching, or peeking behind content for different perspectives on this (see Figure 4).

Information Level and Task Transitions: this application includes seven hierarchical levels of information. The transition between information levels takes place when the user gazes at an element of the tree for 1.3 seconds (defined after pilot tests with three members of the research team). This allows the user to perceive basic properties of the element (e.g., position, name) before the system unfolds it. This timer does not reset immediately when the user looks away, e.g., to briefly explore the surrounding elements. Further, to be lenient in case of potential gaze inaccuracy and make it easy to transition to the next information level, we employ a flashlight technique that considers any elements within a 10°cone of participants' gaze [57]. No selection tasks are supported, and thus no tasks transitions.

4.2.1. Evaluation

We conducted an informal evaluation to get insights into the user experience and usability of the application with the same 12 participants of the previous application. The procedure involved a short briefing about the application and filling out a demographics and consent form. Users started the interaction after the eye-tracker calibration. The goal was to find a path from the origin of life to the humans (five information levels), but participants were informed they were free to explore other elements of the tree as their performance was not being assessed. Each participant interacted for approx. 10 minutes. We relied on think-aloud and observational methods, and followed the task with a brief usability questionnaire and interview.

4.2.2. Results

User Experience. Overall participants were positive about the experience, as they found it "very interesting" (P3, P7, P8, P9, P11) and a "pretty cool use case" (P5, P6, P12) – these were not participants experiencing AR or eye-tracking for the first time, thus we would not consider this a novelty effect (see demographic information).

Scenarios. Considering the unfolding node approach, users found this to be highly useful, as they enjoyed exploring how the tree "evolves" (P7, P9); and found it "well suited for learning" (P6, P9, P11).

Transition. Supporting our goal of enabling implicit explorations of complex 3D visualizations, participants reported that the interaction was "self explanatory" and "intuitive" (P6, P10, P12). One participant forgot the eyes were the main medium for interaction (P10). A few participants suggested that the unfolding activation times did not fit their preferences (P1, P4, P5, P7), indicating that a unified parameter might not be ideal in this use case. Others suggested more user feedback (e.g., displaying a gaze cursor) and control – P3 and P5. This feedback can potentially be explained by several participants' concerns with eye-tracking inaccuracies causing additional effort (especially while walking around the model).

4.2.3. Discussion

Our application was carefully designed to incorporate various information levels. The evaluation was encouraging, with most users being positive about the implicit gaze based traversal

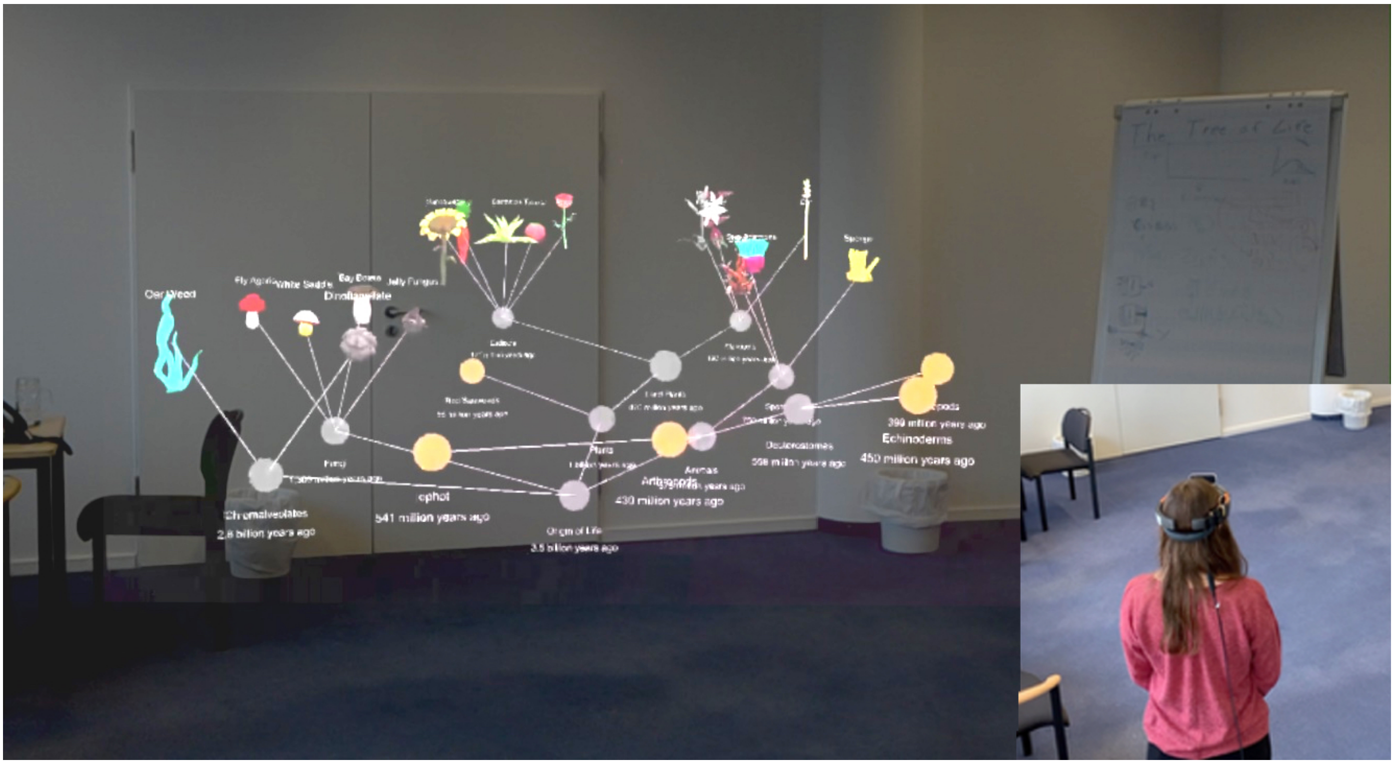

Fig. 8. Tree of Life: gaze provides a way to implicitly navigate between various information levels of a complex 3D visualization of evolutionary biology.

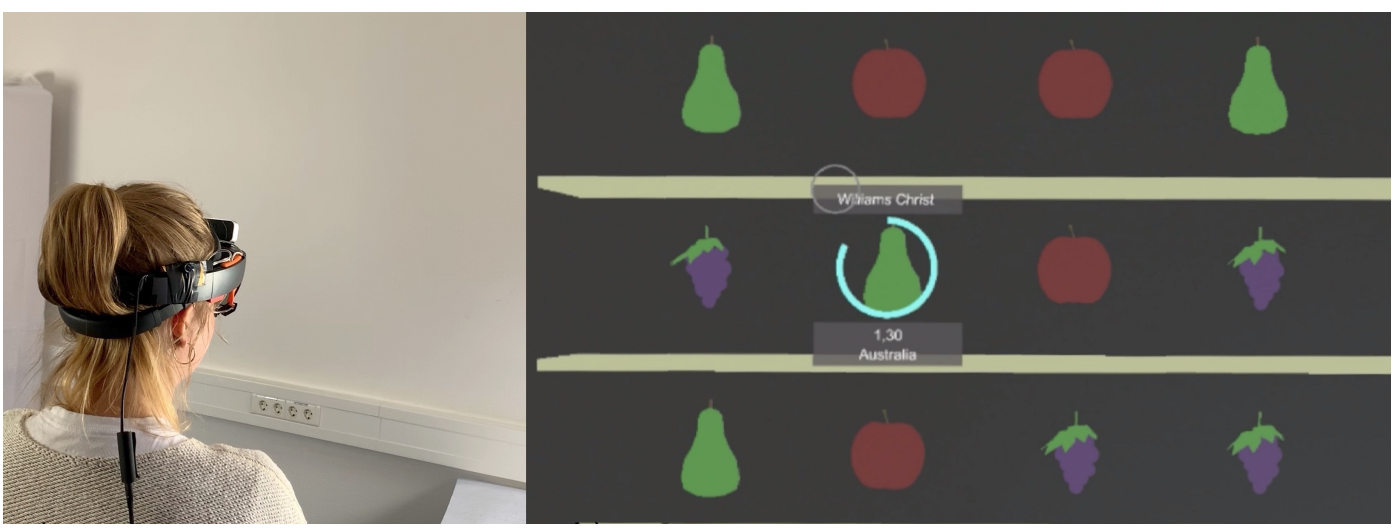

Fig. 9. Shopping: gaze, particularly dwell-time, plays two crucial roles: it can modulate the amount of augmented information displayed in response to users' attention (e.g., show the item cost); and allow users to interact with the item being gazed at (e.g., to "add to cart").

of the various species; this despite the technical accuracy issues that hampered the experience for a few users. Nonetheless, we have observed that gaze is a promising and intuitive tool when designing for gradual, continuous information transitions over time; as opposed to otherwise explicit, more-effort based manual triggers. We believe these preliminary findings to be of use to others considering a wider range of visualizations.

4.3. Application #3: Shopping

Continuum Transitions: our last prototype application represents a more leveled scenario where the user's task (i.e., shopping) takes place as much in the real-world (e.g., navigating a supermarket) as in the context of AR (e.g., checking item information, adding items to a shopping cart). This is representative of various AR scenarios where a visual-search task includes both real and digital elements (e.g., city navigation). Visually, the application presents various shopping items laid out in a vertical 2D grid representing a supermarket shelf. These items represent three classes of fruit (apples, pears, and grapes), and are presented 14 cm away from each other.

Information Level and Transitions: in these scenarios gaze can play two crucial roles (see Figure 9). First, it can modulate

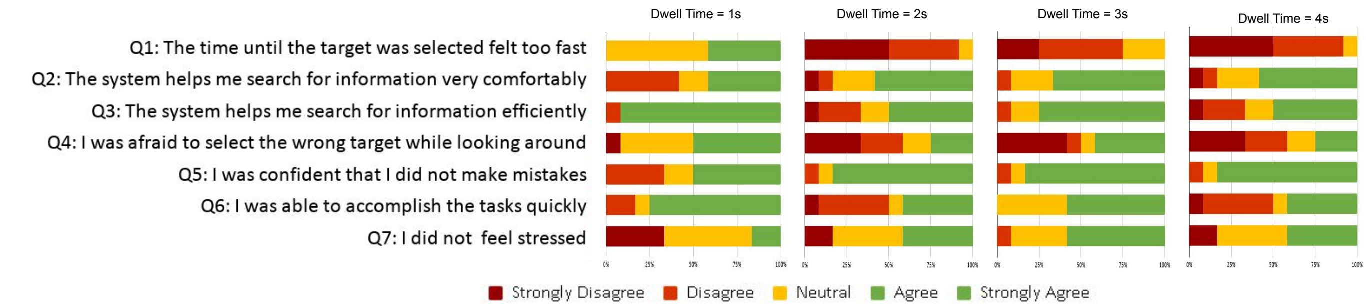

Fig. 10. Questionnaire results for our third prototype application (Shopping). Answers are provided across four dwell implementations: 1, 2, 3 and 4s.

the amount of augmented information displayed in response to users' attention – information level transitions. Without this, the visual clutter alone could make most of these AR experiences unpleasant and ineffective. Second, it supports task transitions by allowing users to interact with the point-of-interest being gazed at via UI selections. The challenge with this approach resides precisely in effectively supporting both types of transitions. In the context of our application, the former supports a single layer of information with the type, price, and country of origin of the item being gazed-at, which fades-in after a 500 ms dwell time. The latter supports "add to cart" functionality after an additional dwell time which we study below, represented as a circular loading bar around the point-of-interest (see Figure 9). To minimize the effect of eye-tracking inaccuracies, these dwell times do not reset when the user looks away, but instead countdown towards zero. Finally, as with the previous application, we use the flashlight technique [57] to aid gaze pointing.

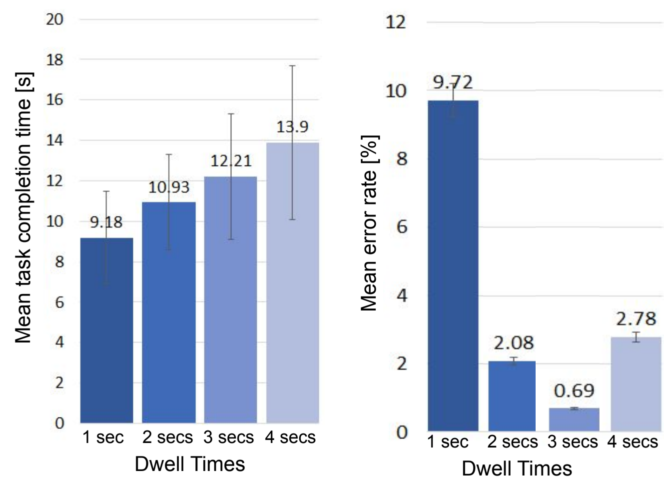

Fig. 11. Task completion time (a) and error rate (b) for the Shopping prototype across dwell implementations (1, 2, 3, 4s). Errors Bars 95% CI.

4.3.1. Evaluation

While various dwell time studies exist, our focus is to better understand how can dwell support both information level transitions and UI selections in a single context. As such, we tested four dwell thresholds for our "add to cart" functionality (1, 2, 3, and 4 seconds, counterbalanced using a Latin square). These were selected based on informal pilot tests looking at the tradeoffs between the minimum time required to consume the item information that faded-in after the first dwell time of 500 ms, and false activations (i.e., wrongly adding an item to the shopping cart). We recruited 12 participants (2F), aged between 21 and 41 years (M = 26, SD = 5.6). All participants had a technical background and, using a 3-point Likert scale (higher is better), were asked to rate their experience with AR (M = 1.58, SD = 0.67) and eye-tracking (M = 1.75, SD = 0.62). None of the participants took part in the first study.

Participants were briefed about the system, completed the appropriate demographic and consent forms, and calibrated the eye-tracker. Participants were then instructed to select 12 items per dwell condition, the information for which was displayed after each selection (e.g., "Select an Elstar apple"). Participants started each dwell condition standing two meters away from a physical wall, were they perceived the augmented shopping interface. This allowed them to see the 12-item grid within the HoloLens' FoV without moving their heads. In this study we collected both quantitative and qualitative results. The former included standard performance metrics such as completion times and error rates. The latter included a 5-point Likert scale usability questionnaire at the end of each condition. The questionnaire was inspired on the NASA TLX, particularly questions on comfort and performance, in addition to bespoke questions designed for the research questions at hand. At the end of the study participants were interviewed and ranked the conditions by preference.

4.3.2. Results

User Experience. Participants stated gaze-based interaction to be easy and intuitive (P8, P6, P7, P8, P11), and "a nice way to display information about the object of interest only when needed" (P12), even though several participants reported gaze inaccuracies at times. Finally, participants preferred dwell times of 1 and 2 s (5 and 6 votes, respectively). One participant preferred dwell times of 3 s.

Transition. We analyzed how users cope with transitioning from consuming information to selecting an UI element, focusing on which timing parameter is more usable. The quantitative results can be seen in Figure 11. As expected, participants completed the tasks faster with a faster dwell-time. The error rate is more indicative of how well users coped with this expand role of dwell. We see that a typical dwell-time of 1 second is not sufficient – leading to a substantial amount in errors at 9.72%.

From 2 seconds onward, the error rate remains stable at about 1-3 %, indicating a more usable parameter in this context. This is supported via an ANOVA test with Greenhouse-Geiser correction, which showed a significant main effect of dwell-time on *completion time* ($F_{11,3=5.05}, p = .007$). As expected, pair-wise comparisons with Bonferroni corrections showed that participants performed significantly faster in the 1 s condition compared to 4 s (p = .02). Further, a significant main effect of the *error rates* among dwell-times was found ($F_{11,3=4.98}, p = .021$). Pairwise comparisons revealed no statistically significant differences between conditions.

The findings are supported by answers to the usability questionnaire (Figure 10). For the sake of brevity, only significant differences are reported. A Friedman test revealed statistically significant differences for answers to questions: Q1, $\chi^2(3)$=22.51, p<.001 – between 1 s and 3 s (p=.007) and 1 s and 4 s (p<.001); and Q7, $\chi^2(3)$=10.43, p=.015, but no post-hoc differences were found significant. Most users preferred dwell times of 1 s (five votes) and 2 s (six votes). A Friedman test revealed a statistically significant difference for the question "The time until the target was selected felt too short or too long", $\chi^2(3)$=22.51, p<.001. Users found 1 s (M=3.58 , SD=0.79) statistically faster than 3 s (M=2, SD=0.74, p=.007) and 4 s (M=1.58, SD=0.67, p<.001). For the question "I was able to accomplish the tasks quickly", there was a significant difference between the responses among the different times, $\chi^2(3)$=10.43, p=0.015.

4.3.3. Discussion

Our evaluation showed that participants could easily cope with the dual nature of dwell in the context of our application – i.e., they were able to effectively browse the AR information without a great number of accidental activations. We found that a combination 0.5 (reveal) and 2s (selection) worked particularly well, despite participants' preference being evenly split between dwell selection times of 1 and 2s. We argue this is likely due to the context of the study, where participants are trying to quickly complete the simulated task that is handed to them. We envision that further in-situ studies need to take place in order to optimize these dwell parameters, and that other visual-search task might require slightly different values.

5. Discussion

In this paper, we contribute a design space for gaze-based interaction for everyday AR scenarios. It covers several roles of gaze that have been previously, if at all, considered in separation: revealing and consuming information; traversing layers of information; and interacting with virtual interfaces. A unique aspect of our design space is that it unifies these dimensions in a single generalized model of high descriptive power, as many potential gaze-based AR interfaces can be articulated with this model. Our point is not that we should build all AR interfaces in such a way, but rather that these three dimensions and how they can be combined, have not been fully explored. We follow up with an exploration of prototype applications that show how the design space fosters new use cases, and that allows to investigate new challenges of gaze interaction such as how different dwell-based interaction tasks can be performed in unison. We summarize our key findings for gaze AR UI design:

Exploit Attention for Background Interactions: The Conversational UI probe demonstrates that even for use cases where the virtual content resides mostly in the background, gaze provides a seamless, in-situ method to rapidly access virtual information without detracting users from their main task. This was exemplified in a task where participants conversed, and could reveal, consume, and also conceal the interface at a glance, without interruption of the conversation flow. This is important, as it sheds light on novel use cases for AR not involving constant augmentations of the real world, such as everyday AR.

Exploit Attention for Large-scale Visualization Traversal: The Tree of Life prototype demonstrates that gaze facilitates the exploration of large, complex 3D visualizations for AR. In particular, it allows us to direct our attention to particular ways and paths within visualizations, by gradually providing more information based on our attention; something that has been previously confined to manual interfaces where each activation of a single node could potentially contribute to physical fatigue and thus a poorer user experience – a point where gaze interaction can become particularly advantageous.

Attention Defines Reality-Virtuality Transitions: Typically, the choice of hardware tends to defines at what point in the continuum an application is. However, when the hardware distinction between these head worn devices blurs, it may become unclear how much reality or virtuality is really experienced through our visual perception channel. In that context, it may be more accurate to utilize our attention: how much do we look at the real environment vs. at virtual content? Even within our field-of-view, part of it might be an augmentation, but never looked at. Understanding and measuring the impact of AR devices that support transitions between the real and the virtual is important to adapt applications toward such use.

Leveraging the ARtention Design Space: The goal of each prototype application was to explore almost exclusively a single dimension of the ARtention design space: the Conversational UI focused primarily on real-world experiences with brief transitions to digital content of low granularity; the Tree of Life focused on the opposite, a rich AR experience with high granularity content; and the Shopping prototype focused on transitions between information consumption and selection tasks. With this in mind, the ARtention design space can now be used to both describe current AR prototypes, and help the design process of new types of AR experiences. These will likely describe richer applications that have a broader engagement with all dimensions of the design space.

A classic AR example is the tourist that attempts to navigate an unfamiliar environment. In this scenario, the user might be following step-by-step AR instructions to its destination, but suddenly attend to a recognizable landmark and make a small detour. Here, the user's engagement with real-world and digital content (and transition between them) will vary greatly, depending on various factors such as schedule flexibility, recognition of real-world markers, or recall of past visits. Likewise, this type of AR application can also provide context-aware information about the user's surroundings such as relevant restaurants around lunchtime. The user can choose to ignore or read these at a glance, engage with this content by unfolding further information such as today's specials (information level transitions), or interact with these to issue a takeaway request (task transitions). In sum, understanding how to holistically and seamlessly support these various transitions in a single application will be crucial for a positive user experience, and ARtention is a rich space in which this understanding can be nurtured.

Attention is not (Only) Gaze: Our works focuses almost exclusively on eye gaze as proxy for attention, and as a result the mechanism in which continuum, information level, and task transitions are enabled. That being said, the ARtention design space can also be considered when other forms of attention proxies are deployed, including the more available and affordable head pointing [58], or a sensor fusion approach combining gaze and brain-computer interfaces [59]. Regarding the later, end-user headsets aiming to infer attention or intent via sensor fusion have been slowly growing in popularity, as exemplified by the latest offerings from HP[1], OpenBCI[2], or Emteq[3]. While several of our takeaways regarding dwell timing parameters and UI feedback would need to be explored in the context of novel input, the design space and its ideas on AR transitions are hardware agnostic.

6. Limitations and Future Work

Our work comes with the following limitations that leave room for future explorations. We provide an extensive design space, however in no way we claim that it is complete. With advances in AR technology and its applications, the design space can be extended and improved. Our application prototypes provide insights into scenarios with currently available hardware. The dynamics of the HMD and the available eye-tracker led to tracking issues, which hampered the user experience – in the future, more robust systems need to expand on the studies to further assess the validity of our results. However, most our findings are high-level and useful as presented.

In addition, many of our spatial and temporal factors were implemented based on pilot testing and iterative designs, however might not be necessarily representative of perfect settings or easy to generalize to others. Other parameters may improve user performance or acceptance, especially when extending to other scenarios. At last, while we covered a range of scenarios, there may be others that would have been an equally good fit.

7. Conclusion

Gaze-enhanced user interfaces can become a helpful addition when AR displays are widely available. The classical eye based interaction by the dwell-time approach provides many opportunities to be extended for the design of interaction techniques when considering the wider range of scenarios that pervasive AR devices can bring forth. We explored three dwell time adaptations, that focus on how a user perceives and interacts with the virtual content that is displayed to assist the user's daily tasks: when people transition from attending to the real environment and to the virtual content, transition between concise and detailed information levels, and between information consumption and selection intent. We presented prototype applications and first user studies, that point to novel use cases of gaze interaction, and that were used to explore and further our understanding of the input parameters usable and also overall the user experience of such an interface controlled completely hands-free, only the user's own directing of visual attention. The main design dimensions are encapsulated in our ARtention design space, which provides a first holistic synthesis of gaze input modes in everyday AR cases. This is relevant as the two areas of AR UI and gaze interaction come with their own challenges, and combined can lead to a more advanced interface where actions happen at a glance; but their design can also present various pitfalls. Underlying are many factors, from what scenarios and interaction tasks are supported, to the depth of information possible, up to detailed temporal and spatial characteristics of assessing visual attention in the system. Our experiments with three application not only verifies the particularities of these design factors, but also provides hands-on insights from user evaluations. Gaze has the potential to reveal information about the world without overloading us, and to select control elements without manual effort – both can be applied across many use cases with carefully considering the design factors presented, pointing to more seamless and implicit interaction capabilities for AR environments.

[1] www8.hp.com/us/en/vr/reverb-g2-vr-headset-omnicept-edition.html
[2] galea.co
[3] www.emteqlabs.com

References

[1] Feiner, SK. Augmented reality: A new way of seeing. Scientific American 2002;286(4):48–55.

[2] Feiner, S, Macintyre, B, Seligmann, D. Knowledge-based augmented reality. Commun ACM 1993;36(7):53–62. URL: https://doi.org/10.1145/159544.159587. doi:10.1145/159544.159587.

[3] Lu, F, Davari, S, Lisle, L, Li, Y, Bowman, DA. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; 2020, p. 930–939.

[4] Buxton, B. Integrating the periphery & context: A new model of telematics. In: Graphics Interface. CANADIAN INFORMATION PROCESSING SOCIETY; 1995, p. 239–239.

[5] Vertegaal, R, et al. Attentive user interfaces. Communications of the ACM 2003;46(3):30–33.

[6] Milgram, P, Kishino, F. A Taxonomy of Mixed Reality Visual Displays (Special Issue on Networked Reality). IEICE transactions on information and systems 1994;77(12):1321–1329. URL: http://ci.nii.ac.jp/naid/110003209335/.

[7] Ajanki, A, Billinghurst, M, Gamper, H, Järvenpää, T, Kandemir, M, Kaski, S, et al. An augmented reality interface to contextual information. Virtual Reality 2011;15(2):161–173. URL: https://doi.org/10.1007/s10055-010-0183-5. doi:10.1007/s10055-010-0183-5.

[8] Davari, S, Lu, F, Bowman, DA. Occlusion management techniques for everyday glanceable ar interfaces. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE; 2020, p. 324–330.

[9] McNamara, A, Kabeerdoss, C. Mobile Augmented Reality: Placing Labels Based on Gaze Position. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct). 2016, p. 36–37. doi:10.1109/ISMAR-Adjunct.2016.0033; iSSN: null.

[10] Kim, M, Lee, MK, Dabbish, L. Shop-i: Gaze based Interaction in the Physical World for In-Store Social Shopping Experience. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. CHI EA '15; Seoul, Republic of Korea: Association for Computing Machinery. ISBN 978-1-4503-3146-3; 2015, p. 1253–1258. URL: https://doi.org/10.1145/2702613.2732797. doi:10.1145/2702613.2732797.

[11] Kytö, M, Ens, B, Piumsomboon, T, Lee, GA, Billinghurst, M. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18; New York, NY, USA: ACM. ISBN 978-1-4503-5620-6; 2018, p. 81:1–81:14. URL: http://doi.acm.org/10.1145/3173574.3173655. doi:10.1145/3173574.3173655.

[12] Azuma, R, Baillot, Y, Behringer, R, Feiner, S, Julier, S, MacIntyre, B. Recent advances in augmented reality. IEEE Computer Graphics and Applications 2001;21(6):34–47. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=963459. doi:10.1109/38.963459.

[13] Khamis, M, Alt, F, Bulling, A. Challenges and design space of gaze-enabled public displays. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. UbiComp '16; New York, NY, USA: Association for Computing Machinery. ISBN 9781450344623; 2016, p. 1736–1745. URL: https://doi.org/10.1145/2968219.2968342. doi:10.1145/2968219.2968342.

[14] Rose, E, Breen, D, Ahlers, KH, Crampton, C, Tuceryan, M, Whitaker, R, et al. 25 - Annotating Real-World Objects Using Augmented Reality. In: Earnshaw, R, Vince, J, editors. Computer Graphics. Boston: Academic Press. ISBN 978-0-12-227741-2; 1995, p. 357–370. URL: http://www.sciencedirect.com/science/article/pii/B9780122277412500293. doi:10.1016/B978-0-12-227741-2.50029-3.

[15] Harrigan, JA. Proxemics, kinesics, and gaze. The new handbook of methods in nonverbal behavior research 2005;:137–198.

[16] Feiner, S, Macintyre, B, Seligmann, D. Knowledge-based augmented reality. Communications of the ACM 1993;36(7):53–62. URL: https://doi.org/10.1145/159544.159587. doi:10.1145/159544.159587.

[17] Julier, S, Lanzagorta, M, Baillot, Y, Rosenblum, L, Feiner, S, Hollerer, T, et al. Information filtering for mobile augmented reality. In: Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000). 2000, p. 3–11. doi:10.1109/ISAR.2000.880917; iSSN: null.

[18] Feiner, S, MacIntyre, B, Haupt, M, Solomon, E. Windows on the world: 2d windows for 3d augmented reality. In: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology. UIST '93; New York, NY, USA: ACM. ISBN 0-89791-628-X; 1993, p. 145–155. URL: http://doi.acm.org/10.1145/168642.168657. doi:10.1145/168642.168657.

[19] D. Lindlbauer A. Feit, OH. Context-aware online adaptation of mixed reality interfaces. In: Proceedings of the 32th Annual ACM Symposium on User Interface Software and Technology. UIST '19; New York, NY, USA: ACM; 2019, p. 213–222.

[20] Zhu, J, Ong, S, Nee, A. A context-aware augmented reality assisted maintenance system. International Journal of Computer Integrated Manufacturing 2015;28(2):213–225. URL: https://doi.org/10.1080/0951192X.2013.874589. doi:10.1080/0951192X.2013.874589.

[21] Lages, WS, Bowman, DA. Walking with adaptive augmented reality workspaces: Design and usage patterns. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. IUI '19; New York, NY, USA: Association for Computing Machinery. ISBN 9781450362726; 2019, p. 356–366. URL: https://doi.org/10.1145/3301275.3302278. doi:10.1145/3301275.3302278.

[22] Jorge, J, Anjos, RKD, Silva, R. Dynamic occlusion handling for real-time ar applications. In: The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry. VRCAI '19; New York, NY, USA: Association for Computing Machinery. ISBN 9781450370028; 2019,URL: https://doi.org/10.1145/3359997.3365700. doi:10.1145/3359997.3365700.

[23] Bolt, RA. Gaze-orchestrated dynamic windows. In: Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '81; New York, NY, USA: Association for Computing Machinery. ISBN 0897910451; 1981, p. 109–119. URL: https://doi.org/10.1145/800224.806796. doi:10.1145/800224.806796.

[24] Jacob, RJK. What you look at is what you get: eye movement-based interaction techniques. ACM Press. ISBN 978-0-201-50932-8; 1990, p. 11–18. URL: http://portal.acm.org/citation.cfm?doid=97243.97246. doi:10.1145/97243.97246.

[25] Qvarfordt, P, Zhai, S. Conversing with the user based on eye-gaze patterns. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '05; Portland, Oregon, USA: Association for Computing Machinery. ISBN 978-1-58113-998-3; 2005, p. 221–230. URL: https://doi.org/10.1145/1054972.1055004. doi:10.1145/1054972.1055004.

[26] Ware, C, Mikaelian, HH. An evaluation of an eye tracker as a device for computer input2. In: Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface. CHI '87; New York, NY, USA: Association for Computing Machinery. ISBN 0897912136; 1986, p. 183–188. URL: https://doi.org/10.1145/29933.275627. doi:10.1145/29933.275627.

[27] Zhai, S, Morimoto, C, Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. ACM Press. ISBN 978-0-201-48559-2; 1999, p. 246–253. URL: http://portal.acm.org/citation.cfm?doid=302979.303053. doi:10.1145/302979.303053.

[28] Pfeuffer, K, Alexander, J, Chong, MK, Gellersen, H. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In: Proceedings of the 27th annual ACM symposium on User interface software and technology. UIST '14; Honolulu, Hawaii, USA: Association for Computing Machinery. ISBN 978-1-4503-3069-5; 2014, p. 509–518. URL: https://doi.org/10.1145/2642918.2647397. doi:10.1145/2642918.2647397.

[29] Stellmach, S, Dachselt, R. Look & Touch: Gaze-supported Target Acquisition. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '12; New York, NY, USA: ACM. ISBN 978-1-4503-1015-4; 2012, p. 2981–2990. URL: http://doi.acm.org/10.1145/2207676.2208709. doi:10.1145/2207676.2208709.

[30] Chatterjee, I, Xiao, R, Harrison, C. Gaze+gesture: Expressive, precise and targeted free-space interactions. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ICMI '15; New York, NY, USA: Association for Computing Machinery. ISBN 9781450339124; 2015, p. 131–138. URL: https://doi.org/10.1145/2818346.2820752. doi:10.1145/2818346.2820752.

[31] Piumsomboon, T, Lee, G, Lindeman, RW, Billinghurst, M. Exploring natural eye-gaze-based interaction for immersive virtual reality. In: 2017 IEEE Symposium on 3D User Interfaces (3DUI). 2017, p. 36–39. doi:10.1109/3DUI.2017.7893315.

[32] Pfeuffer, K, Mayer, B, Mardanbegi, D, Gellersen, H. Gaze + pinch interaction in virtual reality. In: Proceedings of the 5th Symposium on Spatial User Interaction. SUI '17; Brighton, United Kingdom: Association for Computing Machinery. ISBN 978-1-4503-5486-8; 2017, p. 99–108. URL: https://doi.org/10.1145/3131277.3132180. doi:10.1145/3131277.3132180.

[33] Pfeuffer, K, Mecke, L, Delgado Rodriguez, S, Hassib, M, Maier, H, Alt, F. Empirical evaluation of gaze-enhanced menus in virtual reality. In: 26th ACM Symposium on Virtual Reality Software and Technology. VRST '20; New York, NY, USA: Association for Computing Machinery. ISBN 9781450376198; 2020,URL: https://doi.org/10.1145/3385956.3418962. doi:10.1145/3385956.3418962.

[34] Hirzle, T, Gugenheimer, J, Geiselhart, F, Bulling, A, Rukzio, E. A design space for gaze interaction on head-mounted displays. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19; New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702; 2019,URL: https://doi.org/10.1145/3290605.3300855. doi:10.1145/3290605.3300855.

[35] Esteves, A, Verweij, D, Suraiya, L, Islam, R, Lee, Y, Oakley, I. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. UIST '17; New York, NY, USA: ACM. ISBN 978-1-4503-4981-9; 2017, p. 167–178. URL: http://doi.acm.org/10.1145/3126594.3126616. doi:10.1145/3126594.3126616.

[36] Park, HM, Lee, SH, Choi, JS. Wearable augmented reality system using gaze interaction. In: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE; 2008, p. 175–176.

[37] McNamara, A, Boyd, K, George, J, Jones, W, Oh, S, Suther, A. Information Placement in Virtual Reality. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2019, p. 1765–1769.

doi:10.1109/VR.2019.8797891; iSSN: 2642-5246.

[38] McNamara, A, Boyd, K, Oh, D, Sharpe, R, Suther, A. Using Eye Tracking to Improve Information Retrieval in Virtual Reality. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). 2018, p. 242–243. doi:10.1109/ISMAR-Adjunct.2018.00076; iSSN: null.

[39] Esteves, A, Shin, Y, Oakley, I. Comparing selection mechanisms for gaze input techniques in head-mounted displays. International Journal of Human-Computer Studies 2020;139:102414.

[40] Fitzmaurice, GW. Situated information spaces and spatially aware palmtop computers. Communications of the ACM 1993;36(7):39–49. URL: https://doi.org/10.1145/159544.159566. doi:10.1145/159544.159566.

[41] Rekimoto, J. NaviCam:A Magnifying Glass Approach to Augmented Reality. Presence: Teleoperators and Virtual Environments 1997;6(4):399–412. URL: https://doi.org/10.1162/pres.1997.6.4.399. doi:10.1162/pres.1997.6.4.399.

[42] Rekimoto, J, Nagao, K. The world through the computer: computer augmented interaction with real world environments. In: Proceedings of the 8th annual ACM symposium on User interface and software technology. UIST '95; Pittsburgh, Pennsylvania, USA: Association for Computing Machinery. ISBN 978-0-89791-709-4; 1995, p. 29–36. URL: https://doi.org/10.1145/215585.215639. doi:10.1145/215585.215639.

[43] Feiner, S, MacIntyre, B, Höllerer, T, Webster, A. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. Personal Technologies 1997;1(4):208–217. URL: http://link.springer.com/10.1007/BF01682023. doi:10.1007/BF01682023.

[44] Park, HM, Seok Han Lee, , Jong Soo Choi, . Wearable augmented reality system using gaze interaction. IEEE. ISBN 978-1-4244-2840-3; 2008, p. 175–176. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4637353. doi:10.1109/ISMAR.2008.4637353.

[45] vrview: Library for embedding immersive media into traditional websites. 2018. URL: https://github.com/googlevr/vrview; original-date: 2015-10-13T01:45:58Z.

[46] Bernardos, AM, Gómez, D, Casar, JR. A Comparison of Head Pose and Deictic Pointing Interaction Methods for Smart Environments. International Journal of Human–Computer Interaction 2016;32(4):325–351. URL: https://doi.org/10.1080/10447318.2016.1142054. doi:10.1080/10447318.2016.1142054.

[47] Miniotas, D, Špakov, O, Tugoy, I, MacKenzie, IS. Speech-augmented Eye Gaze Interaction with Small Closely Spaced Targets. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications. ETRA '06; New York, NY, USA: ACM. ISBN 1-59593-305-0; 2006, p. 67–72. URL: http://doi.acm.org/10.1145/1117309.1117345. doi:10.1145/1117309.1117345.

[48] Sibert, LE, Jacob, RJK. Evaluation of eye gaze interaction. ACM Press. ISBN 978-1-58113-216-8; 2000, p. 281–288. URL: http://portal.acm.org/citation.cfm?doid=332040.332445. doi:10.1145/332040.332445.

[49] Velichkovsky, B, Sprenger, A, Unema, P. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In: Human-Computer Interaction INTERACT '97. IFIP — The International Federation for Information Processing; Springer, Boston, MA. ISBN 978-1-4757-5437-7 978-0-387-35175-9; 1997, p. 509–516. URL: https://link.springer.com/chapter/10.1007/978-0-387-35175-9_77. doi:10.1007/978-0-387-35175-9_77.

[50] Rivu, R, Abdrabou, Y, Pfeuffer, K, Esteves, A, Meitner, S, Alt, F. Stare: Gaze-assisted face-to-face communication in augmented reality. In: ETRA Adjunct. 2020, p. 14–1.

[51] Behzadan, A, Kamat, V. Visualization of construction graphics in outdoor augmented reality. In: Proceedings of the Winter Simulation Conference, 2005. 2005, p. 7 pp.–. doi:10.1109/WSC.2005.1574469; iSSN: 1558-4305.

[52] Hedley, NR, Billinghurst, M, Postner, L, May, R, Kato, H. Explorations in the Use of Augmented Reality for Geographic Visualization. Presence: Teleoperators and Virtual Environments 2002;11(2):119–133. URL: https://doi.org/10.1162/1054746021470577. doi:10.1162/1054746021470577.

[53] Schall, G, Mendez, E, Kruijff, E, Veas, E, Junghanns, S, Reitinger, B, et al. Handheld Augmented Reality for underground infrastructure visualization. Personal and Ubiquitous Computing 2009;13(4):281–291. URL: https://doi.org/10.1007/s00779-008-0204-5. doi:10.1007/s00779-008-0204-5.

[54] Geller, T. Interactive Tabletop Exhibits in Museums and Galleries. IEEE Computer Graphics and Applications 2006;26(5):6–11. doi:10.1109/MCG.2006.111.

[55] Horn, M, Atrash Leong, Z, Block, F, Diamond, J, Evans, EM, Phillips, B, et al. Of BATs and APEs: an interactive tabletop game for natural history museums. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '12; Austin, Texas, USA: Association for Computing Machinery. ISBN 978-1-4503-1015-4; 2012, p. 2059–2068. URL: https://doi.org/10.1145/2207676.2208355. doi:10.1145/2207676.2208355.

[56] Hornecker, E. "I don't understand it either, but it is cool" - visitor interactions with a multi-touch table in a museum. In: 2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems. 2008, p. 113–120. doi:10.1109/TABLETOP.2008.4660193; iSSN: null.

[57] Liang, J, Green, M. JDCAD: A highly interactive 3d modeling system. Computers & Graphics 1994;18(4):499–506. URL: http://www.sciencedirect.com/science/article/pii/0097849394900620. doi:10.1016/0097-8493(94)90062-0.

[58] Özacar, K, Hincapié-Ramos, JD, Takashima, K, Kitamura, Y. 3D Selection Techniques for Mobile Augmented Reality Head-Mounted Displays. Interacting with Computers 2016;29(4):579–591. URL: https://doi.org/10.1093/iwc/iww035. doi:10.1093/iwc/iww035. arXiv:https://academic.oup.com/iwc/article-pdf/29/4/579/17709962/i

[59] Putze, F, Weiβ, D, Vortmann, LM, Schultz, T. Augmented reality interface for smart home control using ssvep-bci and eye gaze. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE; 2019, p. 2812–2817.