Guidelines for Productivity in Virtual Reality

Mar Gonzalez-Franco and **Andrea Colaco,** Google

Insights

- VR headsets can become productivity tools if we enable multitasking and transitions in and out of devices.
- Inside VR, people can achieve higher focus and improve remote collaboration.
- New combinations of multimodal input will need to enable fast and high-precision work in reachable and unreachable spaces.

IMAGE BY HEMIN XYLAN / SHUTTERSTOCK.COM

Most of our interactions with digital content currently occur inside 2D screens. Moving from that format to immersive setups, however, brings a paradigm shift: from content inside the screen to users inside the content. This change requires us to revisit how we blend the analog and the digital and how we transfer content between the two modes—perhaps it even asks for new guidelines, too. While different solutions appear in the space, the gulf between the two worlds only seems to widen. We can start to see what works, and what does not work so well, in an empirical or ethnographic approach, beyond laboratory studies. But if we want to accelerate adoption, we need to better understand how current tasks can be improved and how this new form of interaction can increase productivity. In this article, we analyze and converge what we think works, and envision how this new set of immersive devices and interactions can enable productivity beyond already existing tools.

BACKGROUND

We'll start with some lessons from previous inflection points in computing using a simplified history of events. First, we had computers. They became smaller and cheaper so they could be used at home. Then, the Internet

connected everything. Tech then got even smaller, so much so that we could move to mobile computing, which eventually jumped to our phones. And that is where we are now.

In retrospect, we can see how all these evolutionary leaps have coexisted. Despite all sorts of predictions, phones didn't kill the PC, and probably neither will immersive tech. They are likely to coexist for the foreseeable future. This is one of the early points we want to make: the importance of interoperability.

Let's consider a world where digital content finally moves out of 2D screens into our 3D worlds (real worlds). As opposed to having two realities—a metaverse (virtual) and a real world—extended reality (XR) aims to blend the two. In this framework, there are some things that will continue being 2D planes even inside 3D; for example, documents. But these panels might resynthesize in our surroundings in more-affordable ways than on an actual screen: on our walls, on top of tables, attached to other objects that make them tangible, or taking into account the user for optimal ergonomic size and position. And yes, that means sometimes content might be just shown as screens inside XR. But there will always be a component of spatially arranging things, whether they are 2D or 3D.

Let's take a deeper look at the spatial and ergonomic components together. We will focus on wearable sets of glasses (even in the primitive form of an AR phone; since it is held in the hand, it could be temporarily considered a wearable). This type of immersive tech is an interface between the user and their environment, inviting embodied interaction. The dichotomy is then between body-locked content and interactions versus world-locked ones. Interestingly, in XR the boundaries are more dynamic; our input systems as well as our content will have many more options, which we discuss later.

Mundane productivity topics, like interoperability or input, are the scaffolding for collaboration, multitasking, transitions, and interruptions. There are many other things users will need to do, but these are the ones we focus on here and that we believe are core for productivity.

If these topics are not addressed well, VR will not be widely adopted for work scenarios, especially for information workers, who currently spend most of their time using PCs. *What can VR do better?* is a good research question. For now, though, let's make sure VR isn't worse at these topics than current PCs.

INFORMATION WORKERS

Information workers, also known as knowledge workers, are those who spend most of their productive time enabled by computers. Other workers, such as frontline workers in factories, farms, or other real-world settings, will also experience a big improvement in productivity when they are able to augment their realities. In fact, the impact on their productivity will perhaps be even greater than that of information workers (Figure 1), as many real-world tasks still lack advanced assisted computation, whereas the improvement in productivity with VR for information workers might be marginal.

In this article we focus on how immersive tech will revolutionize productivity for information workers. This narrower initial focus is for three reasons:

- Information workers are already intensively using devices and adopting new software and gadgets on a regular basis. They can be considered early adopters, the power users of digital content.
- Safety issues will be reduced. Working at a desk is a much safer control space. Locomotion, the need to move around, is reduced and scenes are more constrained, with more-limited sets of objects, reducing dependencies from scene-understanding algorithms. It is an ideal petri dish for early XR, where the real world can start to blend with VR, with pass-through views and so on.
- Ergonomic issues will be reduced. Head-mounted displays (HMDs) in a multidevice scenario might be used only temporarily, which means that tethered cables, as well as the size or weight of the device, might be less critical.

It's clear that HMDs, computer vision, and AI will improve over time, enabling many other forms of XR. This follows the trend we have already seen with other specs that have improved in the past decade. We can comfortably read text in most HMDs, with 4K displays complemented by advances in optics—pancake lenses or three-element lenses—finally making these devices viable for information workers.

WHY ADOPT VR FOR PRODUCTIVITY?

If people are going to adopt VR for work, it is because some tasks become easier or faster to perform in this medium. Researchers have been trying to find which specific tasks benefit. We can, for example, augment experiences by visualizing more information in context, augment presentations inside VR, and improve meeting experiences.

Indeed, meeting with other humans is an experience complex enough that we haven't managed to re-create it with video conferences in 2D; maybe VR can help with its spatial audio and vision. VR can enable more ecological validity and unlock evolutionary wonders such as directed attention, peripersonal spaces, and concurrent taking, as well as unlock the ability to use our body for interaction (pointing, gazing, and enabling spatial formations). Perhaps it will even enable the use of whiteboards in direct ways by multiple users, and ultimately support a form of collaborative spatial work unavailable with traditional 2D screens.

In general, there is agreement that experiential tasks are good candidates to be improved with VR, even if they don't require colocated participation. That means going beyond meetings, to affect learning with improved recall and hippocampal activity. These experiential tasks are particular use cases for productivity, however, and might not justify full adoption of VR.

If these topics are not addressed well, VR will not be widely adopted for work scenarios, especially for information workers, who currently spend most of their time using PCs.

A different look into VR productivity opportunities can focus on the uniqueness of the medium, instead of on specific tasks. VR has traditionally been labeled as an isolating medium. But despite the fact that this isolating effect can go away with current and future video pass-through technology, the ability to transform HMDs into a monastery on demand has big potential for helping with productivity. That could mean XR offers an increased capacity to upstream focus, creating fewer distractions so we can channel larger chunks of attention to work. The monastery example is perhaps very extreme, but it is illustrative of the "private space on the go" potential of this technology.

Even if it sounds like the antithesis of focus, the other superpower of VR could be its scaffolds for multitasking. These scaffolds would be enabled by its large horizontal field of view (FOV), which would become the largest real estate display of any available device, providing users access to an optimal set of concurrent screens, applications, and fast layouts at the same time. These two properties transcend the type of task and highlight particularities of the medium. If you want to focus, go to VR. If you want to multitask, go to VR. These will need to be enabled by both the hardware and software, however, and with good practices, which we highlight below.

PASS-THROUGH VERSUS SEE-THROUGH

In pass-through video HMDs (Figure 2), a complete occlusion of the real world is possible by turning off the video feed. A set of cameras record the real world and then stream the recording to the (opaque) displays inside the headset. When the camera feed is off, a user feels they are in another location entirely—full VR—and their presence in that location can be so strong that they forget about their real-world surroundings [1].

In optical see-through HMDs, the user wears transparent glasses and can overlay projections of synthetic content on top of the real world (Figure 2). In see-through devices, totally occluding the world would probably require a display the size of the human FOV, and the technology isn't there yet.

But the uninterrupted work-focus scenario—independent, if one is writing a document on a screen, answering emails, or preparing a set of slides—also needs to be compatible with the times when users will want to multitask and be connected to other devices and their real world. Most people don't want to live in isolation for their entire work life. They need to be able to transition in and out of such an immersive device, with reduced mental-switch burden.

Figure 1. Vignettes showing use cases of immersive technology for productivity. Left: Complementing the information worker experience. Center: Augmenting the real world for frontline workers, with in-context access to information (e.g., with instructions on fixing a broken device). Right: Enabling a factory worker to better operate, design, and control a process through the use of augmented reality tools.

Figure 2. Two main ways of blending digital and real content, with either optical see-through or video pass-through (sometimes referred to as video see-through). In optical see-through, the digital world is projected on a surface that has a level of transparency. In video pass-through, the eyes are completely occluded from the world with opaque displays.

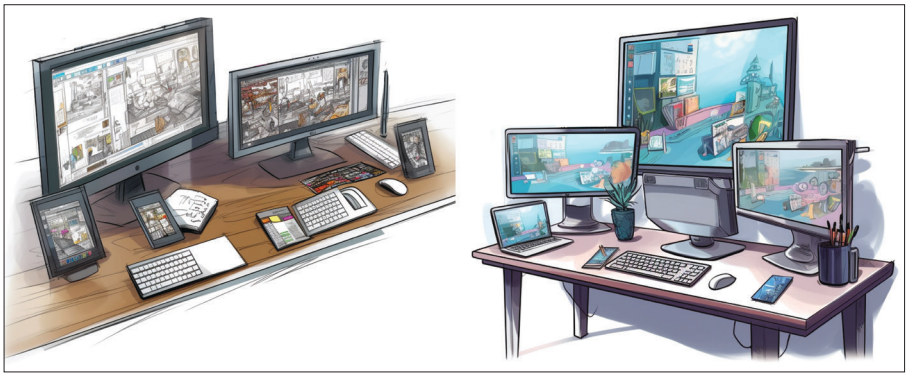

Figure 3. Sketches of real desktops with multiple displays and devices.

GUIDELINES

An HMD can become a complete interface between the user—their body and brain—and their environment. As a wearable, it is continuously adapting in first-person perspective—looking out, projecting in—affording an opportunity for reimagining computing. But we also need to make sure the basics are covered.

At the bare minimum an HMD should do very well what the other devices already do: interacting with digital content. So, even at the risk of sounding trite, here we highlight the need to make sure this technology is interoperable, has good input systems, mediates interruptions and transitions, is accessible, and allows for multitasking. And that it does all this while reducing the mental cost of switching in and out. Perhaps, then, the

Figure 4. Architectural layout of two rooms with the overlay of common free space (red stripes: unavailable; green: free). In most cases, and as the number of users increases, areas of available overlay on dissimilar spaces will tend toward zero.

"killer app" for HMDs is just a very good interaction paradigm that simplifies the use of this technology on a daily basis, even if for very short periods of time.

Interoperability. Introducing any new device to an ecosystem comes at a cost. Immersive technologies for information workers arrive in a space that is already heavily populated with other devices. Workers use a large set of layouts with multiple devices and display configurations in their offices (Figure 3). They might be in a semipermanent setting or on the go, on mobile devices, tablets, or laptops. Workers expect their multiple devices to transfer content seamlessly and to be able to use the same set of apps with corresponding actions, perhaps even with the same inputs. This is a key requirement for VR devices that need to be designed within this context: to account for solo users who transition quickly between devices.

Collaboration. Interoperability is not just a single-user issue. People work together. VR users cannot expect everyone to be wearing an HMD. This will be especially true for early adopters, who will face hybrid interaction paradigms when other users don't have VR headsets. This puts emphasis on the importance of figuring out ways both for traditional users to engage in VR collaborative spaces and for VR users to appear on their collaborators' 2D tools. For example, using avatars to represent VR users might make more sense inside a regular videoconferencing tool [2] than inside collaborative VR, where the focus should instead be on how to tile spatially correct 2D participants in coherent spots of the VR environment.

Even as adoption grows, when more people have HMDs, users will not be able to assume their current spaces are similar enough to have totally free collaborative environments (Figure 4). It will be hard to share immersive spaces between people, and interaction might need abstractions of semantics from motions, meaning that if you are, for example, pointing at one object in your environment but that same object is positioned in a different location in the other person's environment, we will have to adjust that interaction. There will be some artificial repositioning of users and content in space according to the scene understanding in each specific case, trying to maintain certain interaction consistencies that enable both communication and collaboration.

Interruptions and transitions. Interruptions and transitions have a significant impact on productivity and workflow. Coworkers, kids, pets, app notifications, other devices, calls, emails, messages—they can all disrupt focus, break momentum, and lead to inefficiencies. While transitions between contexts to respond to interruptions isn't just a problem of immersive setups, it can be amplified in VR, where the HMD creates a visual barrier to the external world that can be overcome only with a good system for detection and mediation of interruptions.

The truth is that inside VR even ordinary activities like drinking coffee could be considered interruptions. Being able to access the real world and bring parts of it to the VR environment in a blended manner will be key. This transformation of an isolated VR experience into an extended reality (XR) that is connected to its surroundings is essential for effective productivity.

Once interruptions are presented inside the HMD, users will expect to seamlessly transition from one activity to another, perhaps even in and out of their devices.

One effective strategy is to handle as many external interruptions as possible inside VR, minimizing the need to take

This is a key requirement for VR devices that need to be designed within this context: to account for solo users who transition quickly between devices.

off the headset. That means VR systems should ensure other devices are tracked, visible, and interactable, and that their screens are mirrored inside VR. Smart pass-through allows users to view the real world without taking off the headset. But full pass-through might not always be necessary. There are many different forms of adaptive pass-through that can enable this XR experience, ranging from full pass-through, segmented objects, and digital versions of the real world reconstructed via Gaussian splats or neural radiance fields (NeRFs). Dynamic chaperones, for example, can display relevant information through edge-detection filters that activate when movement is detected near the boundaries, serving as an intermediate step before enabling full pass-through. Partial pass-through and segmentation techniques can also be employed to preserve a sense of connection with real-world landmarks, such as a sofa, window, or bed, while immersed in VR—a sense of presence in the real world.

Indeed, presence is a unique feature of VR technologies that can create the illusion of being elsewhere, in another place; for users, this often means losing track of their real space [1]. In terms of productivity, however, it may be desirable to enable users to feel present simultaneously in both the virtual and real spaces.

One solution to be in two places—the virtual and the real—at the same time is to bring users to an intermediate space that closely resembles their current physical environment but in an improved format: a clutter-free, clean, and productive setting. By curating objects through blended reality, VR can effectively "clean your room," eliminating distractions and aiding concentration. Leveraging generative AI tools, XR can go beyond simply removing elements from the scene and generate an entirely new space where users feel present in both the physical and virtual realms (Figure 5).

If this mediated interruptions system is well managed, it can reduce the likelihood of unnecessary interruptions while facilitating transitions and help achieve focus.

Multitasking. Multitasking is an essential skill that greatly affects productivity, and it is closely tied to seamless transitions. Ultimately, multitasking involves handling activities simultaneously while swiftly switching between multiple tasks. It can be considered a by-product of an effective interruption management system. While some argue that multitasking can decrease productivity, when used appropriately it can increase efficiency, enabling individuals to make progress on multiple tasks concurrently. Multitasking allows people to optimize their time by avoiding idle periods; for example, when downloading files or waiting for a computational response, one can work on another task, thereby maximizing productivity throughout the day. Additionally, multitasking can help prevent monotony and provide mental stimulation. Though it might seem counterintuitive, switching between tasks helps individuals maintain focus and interest, effectively combating boredom and fatigue. This, in turn, can promote higher levels of engagement and motivation.

In XR, multitasking can be further enhanced by the wide horizontal FOV and head tracking, which in essence create an "infinite 360" real estate around the user that allows for multiple display arrangements. The ability to handle multiple tasks simultaneously and change layouts is also very inviting to render other devices inside the HMD, so the user can have everything accessible in one place.

Input system and content. Input is perhaps the hardest issue for VR, with numerous new interaction paradigms and input combinations to explore and an ever-divergent vocabulary of actions and interactions that continues to expand (Figures 6 and 7). HMDs offer new ways to interpret intent, attention, and action through enhanced sensing capabilities and wearable formats. But these newfound capabilities also come with challenges in terms of expressiveness, and often input

Figure 5. Extended reality (XR) can also be used to declutter spaces or even transform spaces using generative AI to become less cognitively demanding (GenXR), where it's easier to focus. Some basic aspects remain, however, like the layout or the interactable devices to facilitate context switching. In this figure, we showcase how the decluttering can be enabled in pass-through situations, by inpainting (top) or by using generative AI that renders a whole new space with similar structural constrains (bottom).

methods that capture the intent or the action might be lacking in other ways and not scale well to all activities.

Reachability versus vision ergonomy. Embodied interaction is finally a possibility, and many have been enamored of *Minority Report* paradigms, using hands to reach virtual objects and grab them (Figure 6). The problem is twofold; not only can we not assume everything will be within reach, especially for individuals with accessibility needs, but also bringing content too close can be visually uncomfortable, as it strains vergence and accommodation (it is un-ergonomic, for example, to look at things very close to the eyes) (Figure 6). The area within 30 centimeters from the eyes can be considered a no-zone for those reasons [3]. At the same time, users' arm length is also limited, working best at two-thirds of its extension. Therefore, there is a small space, roughly between 30 to 50 centimeters, where interfaces will be comfortable for both reach and vision. This highlights the reality that trade-offs are necessary. Objects beyond the reachable space, typically considered to be more than 70 centimeters, require alternative forms of interaction, such as remapped motions, pointing, and ray casting from the head, hands, or eyes, with additional tracking, dwelling, gestures, or combinations thereof.

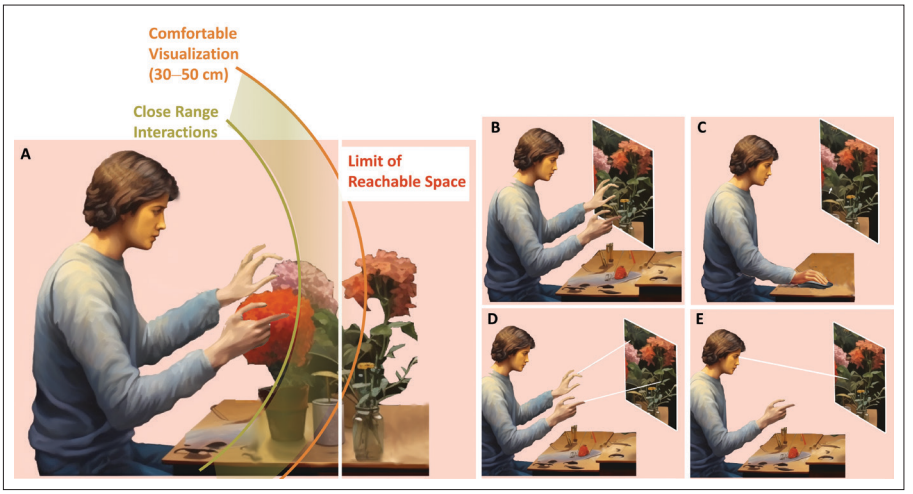

Figure 6. A) Ergonomic spaces around a person inside VR are primarily determined by the comfortable visualization range. B) Within the intersection of reachable space and comfortable visualization, close-range interactions with the body and direct manipulations of the content are possible. C) Remapping of motions to a cursor could be used for interacting with content that is beyond reachable space and/or to reduce fatigue. D,E) Ray casting from the hands, eyes, and/or head can serve as input modalities for far-off content.

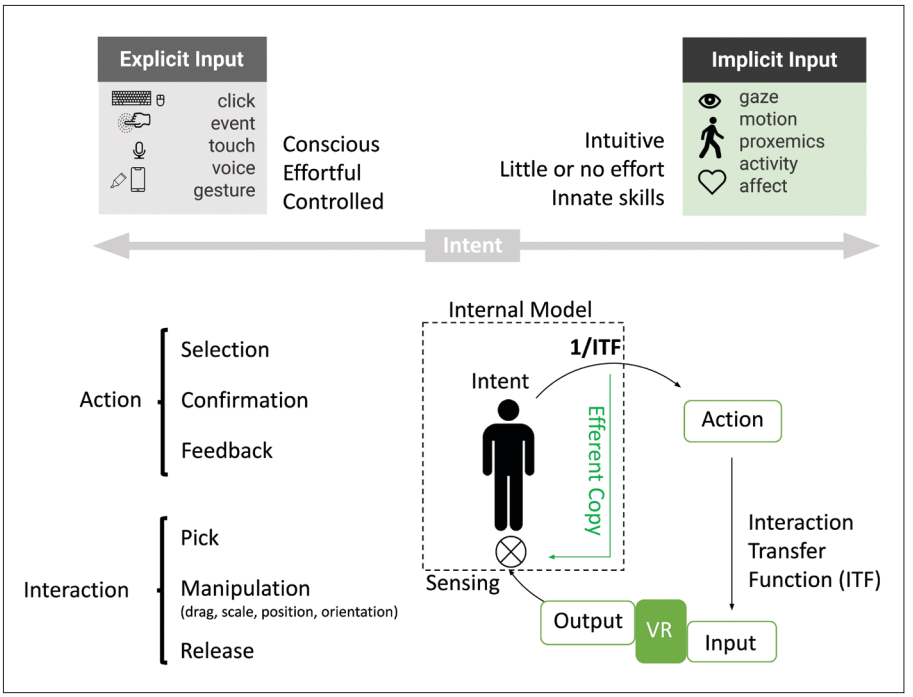

Figure 7. Classification of types of input by intent. The additional complexity is that for some cases the same input can be explicit or implicit depending on whether it has been done consciously or not (e.g., gestures) [6,7]. Different user actions can be aggregated from a series of micro-operations conforming to particular interactions. On the loop we simplify a schematic of the human model of motor control [4] that allows for learning of the interaction paradigm, optimized by the transfer function.

Moreover, users can manipulate the virtual world to reach the unreachable, enabling direct interaction again. Clutching is a form of temporarily attaching the content to the user's body. However, relying heavily on clutch mechanisms to transition between reach and vision comfort would significantly increase the time required to perform any task, and introduce yet another item on the vocabulary of interactions that users will need to learn.

There are other ways to bridge the reachability gap without resorting to clutching. Interaction at a distance can be achieved through ray casting or remapping by abstraction from a gesture, gaze, or posture. These abstractions from implicit inputs, however, can make them prone to unintended interactions [4].

Body locked versus world locked. When rendering content, a key question arises: Is the content attached to the person? Generally, the ability to track and adjust the content position relative to the user opens the possibility of having both body-locked and world-locked content. Transitions between modes, such as clutching to extend reach, become possible.

A starting recommendation would be to match existing affordances of the real world: Assume content will be world locked and input interaction body locked. In the real world, we interact with surrounding objects, such as grabbing a cup of coffee, and at that moment the object becomes body locked—it has been "clutched" (Figure 7).

This approach will provide better visualization ergonomics and supports better mental-mapping persistence, allowing users to remember where they placed something, like a spatial anchor.

There are perhaps some exceptions, such as notifications or menus, or cases involving extreme distances that benefit from proximity to the user. In general, users can either have inputs that work from a reasonable distance, bring the content closer for direct manipulation, or employ locomotion to interact with the content.

Interaction paradigms. The input process can be more precisely explained when abstracted into actions and interactions. Actions can be further divided into micro-operations: selection, confirmation, and feedback [5], while the interactions provide the events with meaning retrospectively, such as manipulation, pick, and release, and will create a vocabulary of user actions together with the context (Figure 7). Actions are more basic primitives that lack the semantics of the larger task.

Mastering any input system means creating a good internal model of this vocabulary. Without a model, users have to rely on high cognitive processing of the sensorimotor feedback all the time. This has a cost. Feedback-monitoring loops for driven actions are rather slow (400 milliseconds) when compared with internal models of motor control (100 milliseconds) in the brain [4], with its corresponding impact on reaction time and increased errors.

In practice, this means that for humans to be able to learn, the process needs to be deterministic. Additionally, a good interaction paradigm should aim to create an easy vocabulary of actions and context combinations that can be internalized to achieve expertise.

One key aspect to create a deterministic system is to minimize false positives. For that, confirmation actions can be linked to more-reliable intentional explicit inputs—a click, a pinch, a particular voice command—while selection can be more blended with implicit and explicit inputs such as pointing and eye gaze. If the same implicit input needs to be used as both selection and confirmation, then the best option generally is to use a dwell timer as confirmation. With good awareness of the environment, context can also be used to improve intent.

Let's consider one example of suboptimal and optimal interaction paradigms: hand interactions. If selection is unreliable due to occlusions, jitter, or a non-negligible error rate in the tracking system, and if gesture recognition for confirmation also introduces additional error rates, the result is a suboptimal interaction system.

Now let's introduce this system as an alternative to someone who regularly uses a mouse for eight hours a day—a high-precision tool that has been optimized and has remained stable for almost 30 years, offering just two primary clicks, right and left, to access a whole vocabulary.

Input guidelines. For XR input in productivity scenarios, we suggest the following guidelines, always bearing in mind that the vocabulary of interactions will need to be internalized and learned by the users, so it will need to feel deterministic and easy (Figure 7):

- *Backward compatibility.* Traditional peripherals offer well-known and precise, explicit input techniques. A mouse with depth [8], an augmented physical keyboard, and other trackable devices like phones and tablets can become input tools for the information worker. They are readily available, offer high precision, and are already familiar to users.
- *Embodied interactions.* Midair gestures can be physically tiring and have lower precision. Reliable hand tracking is essential for successful hand gesture input. If the tracking is not reliable enough, hand tracking will still serve as a valuable tool for communication, and as a backup input when a user doesn't have access to others.
- *Combined techniques.* Different combinations of input modalities can work well for specific users and applications. Generally, explicit input methods (e.g., mouse, voice, etc.) can be amplified and complemented by implicit interaction signals like eye gaze or head gaze. While one might be used for selection, the other might work as the confirmation of input.

CONCLUSION

This article explores the potential of virtual reality for enhancing productivity, particularly for frontline workers, while also addressing the challenges that must be overcome. With the proposed set of guidelines, we aim to reduce the friction of transitioning between devices, that is, coming in and out of the VR headset when working in combination with a laptop or desktop PC. Additionally, we aim to simplify and streamline interactions within the XR environment, making them straightforward and predictable, while harnessing the enhanced capacity for focus and multitasking offered by VR headsets.

ENDNOTES

1. Sanchez-Vives, M.V. and Slater, M. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience 6*, 4 (2005), 332–339.
2. Panda, P. et al. AllTogether: Effect of avatars in mixed-modality conferencing environments. *Proc. of 2022 Symposium on Human-Computer Interaction for Work*. ACM, New York, 2022.
3. Shibata, T., Kim, J., Hoffman, D.M., and Banks, M.S. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision 11*, 8 (2011), 11.
4. Padrao, G. et al. Violating body movement semantics: Neural signatures of self-generated and external-generated errors. *Neuroimage 124* (2016), 147–156.
5. LaViola, J.J., Jr., Kruijff, E., McMahan, R.P., Bowman, D., and Poupyrev, I.P. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2017.
6. Schmidt, A. Implicit human computer interaction through context. *Personal Technologies 4* (2000), 191–199.
7. Argelaguet, F. and Andujar, C. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics 37*, 3 (2013), 121–136.
8. Zhou, Q., Fitzmaurice, G., and Anderson, F. In-depth mouse: Integrating desktop mouse into virtual reality. *Proc. of the 2022 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2022.

Mar Gonzalez-Franco is a neuroscientist and computer scientist at Google. Her work is at the intersection of human perception and computer science. In her research, she fosters new forms of interaction that will revolutionize how humans use technologies. Her interest lies in spatial computing and on the wild use of technology.
→ margon@google.com

Andrea Colaco is a software engineer at Google introducing novel applied machine learning techniques for context-based human input and intent understanding into new product categories like AR/VR and connected home devices. With a background in computational techniques and computer vision, she studies how these tools bring real-time systems to the next level.
→ andreacolaco@google.com

DOI: 10.1145/3658407