Vision-Based Technology: Next-Gen Control

DEVELOPED FOR: adeia

Vision-Based Technology: Next-Gen Control

Consumers have rapidly become accustomed to various methods of interaction with technology and devices, such as touch, motion, and voice control. Consumers are most familiar with touch-based interfaces thanks to its widespread use across multiple technology sectors, including mobile devices, automotive infotainment, banking, and computing.

Vision-Based Interfaces

provide hands free control in place of or to supplement other forms of interaction and control input, such as tactile (touch) or vocal (voice) input.

The principle of vision-based interaction (VBI) relies principally upon computer vision (CV).

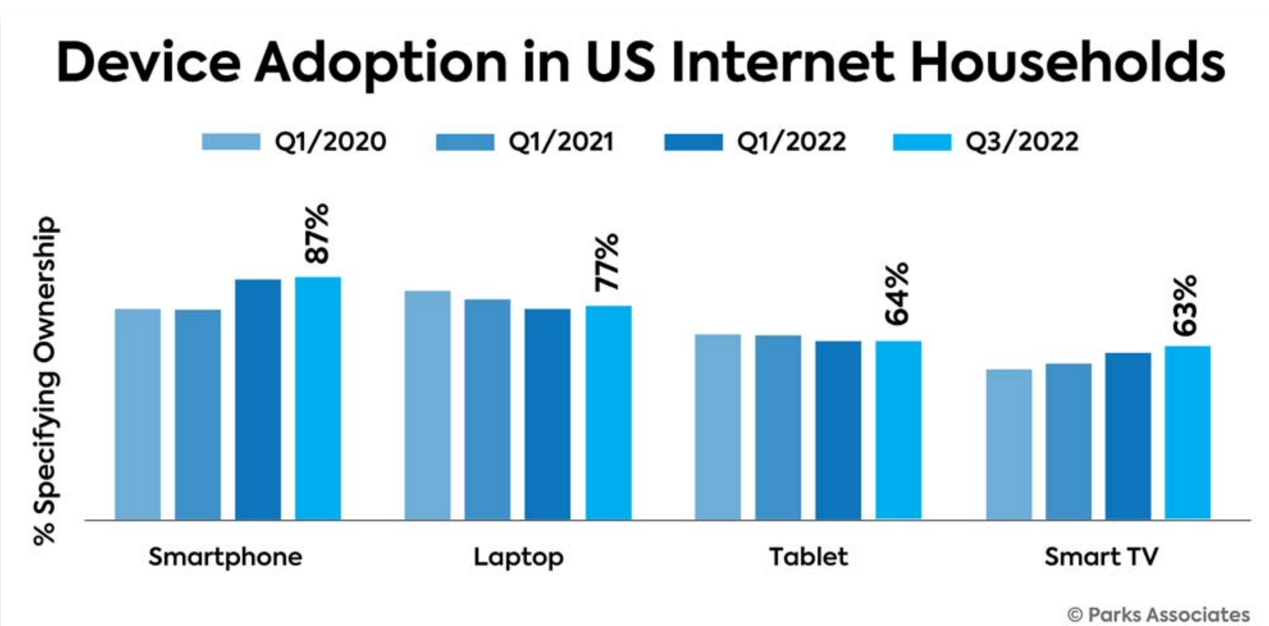
New advancements and technologies are always on the horizon that can advance the interaction paradigm between people and the devices around them. Vision-based interaction is a more recent innovation that can enhance or replace the use of touch-based interfaces to accomplish many of today's tasks, reducing friction in the experience. It can also bring new benefits and features to the millions of people with a disability that limits their ability to use the traditional methods of human-technology interaction.

This whitepaper examines selected consumer technology devices that can potentially be augmented using a vision-based interface, how vision can complement or replace the common touch-based interactions, and the benefits of vision-based interfaces to both consumers and industry. The whitepaper also examines best practices and real-world examples of vision-based interfaces and interactions.

Today's Most Adopted Devices

The potential for vision-based interfaces is incredibly broad across consumer, enterprise, industrial, medical, transportation, military, and other industry verticals. The size of this opportunity can be gleaned from examining one subsegment of applicable consumer devices – everyday technology devices that consumers interact with via touchscreens or tactile input. Popular devices that utilize touch-centric interaction include smartphones, tablets, smart TVs, and laptops.

Currently US households have an average of 16 connected devices, including 11 from the mature CE category, three from smart home, and two from connected health.



adeia PARKS ASSOCIATES

Smartphones

The smartphone is the most common of all touch-based devices and a focal point of human-technology interaction.

- Smartphones are present in 87% of US internet households.

Previous attempts at vision-based interaction in the smartphone industry have been unsuccessful, relying upon either limited computer vision capabilities based on the smartphone's own processor and cameras, eye-tracking on tablets and smartphones, or the use of near-field sensors. Google Glass is the most prominent example of a vision-based product that arrived with lots of hype but ultimately failed. Consumers' lack of interest stemmed from a variety of reasons, including high cost, poor UI/UX design, low battery life, and lack of substantive tangible improvement over legacy touch interaction. Some smartphones continue to use their front cameras for attention tracking, but the industry overall remains focused on touch as the main interaction modality for smartphones.

Tablets

In terms of control and interaction, tablet dynamics are nearly identical to those of smartphones but with even larger display and touch input area. Like smartphones, there is a limited amount of non-touch interaction possible today, primarily via voice assistants.

- Tablets are in 64% of US internet households.

Smart TVs

Smart TV manufacturers are constantly battling the pressure of commoditization, and the control and user interfaces are a constant area of focus and attempted differentiation.

- Smart TVs are now in 63% of internet households.

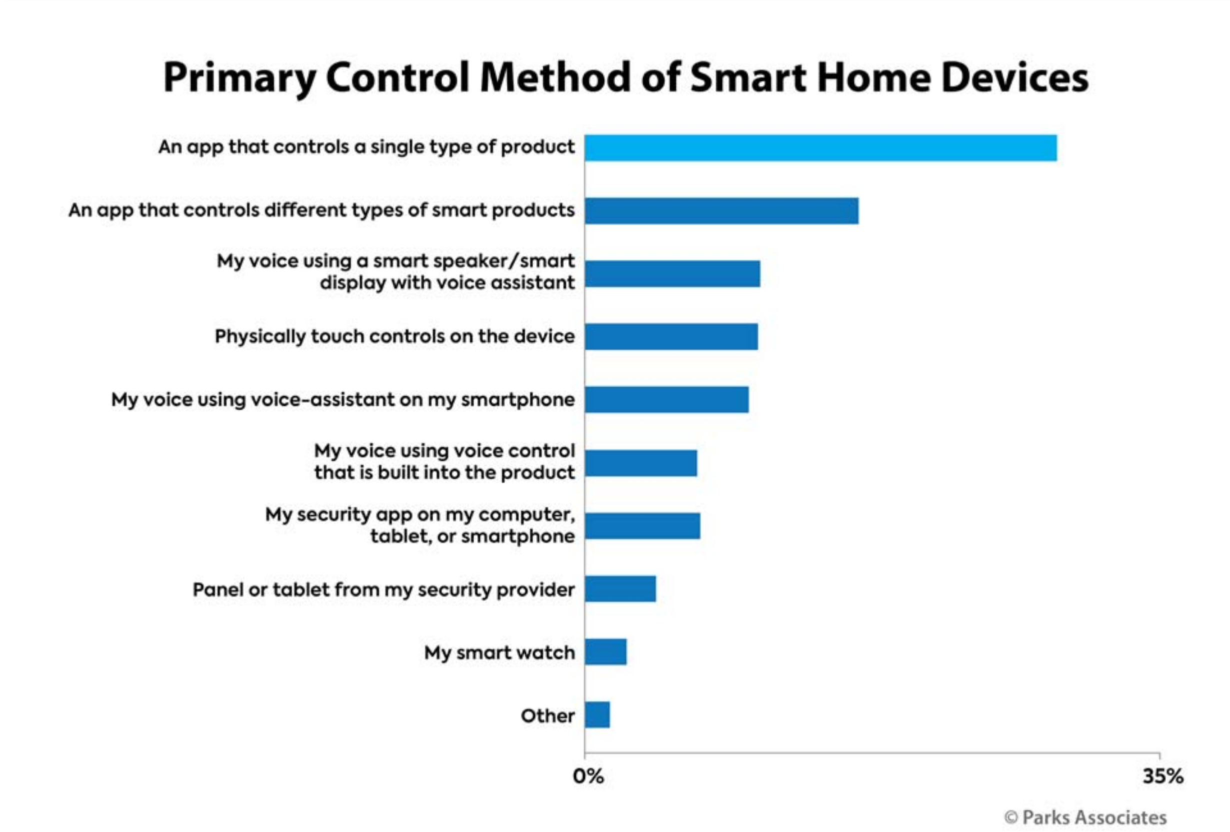
While control of the TV has evolved some, the dominant control method is still through the device's remote control or a remote-control app on a mobile device. Consumers can use voice input to control the TV, through built-in microphones or smart speakers, but this method is merely a supplement to what remains primarily a manual control paradigm. For the top streaming video products—smart TV and streaming media players—voice control is not used frequently.

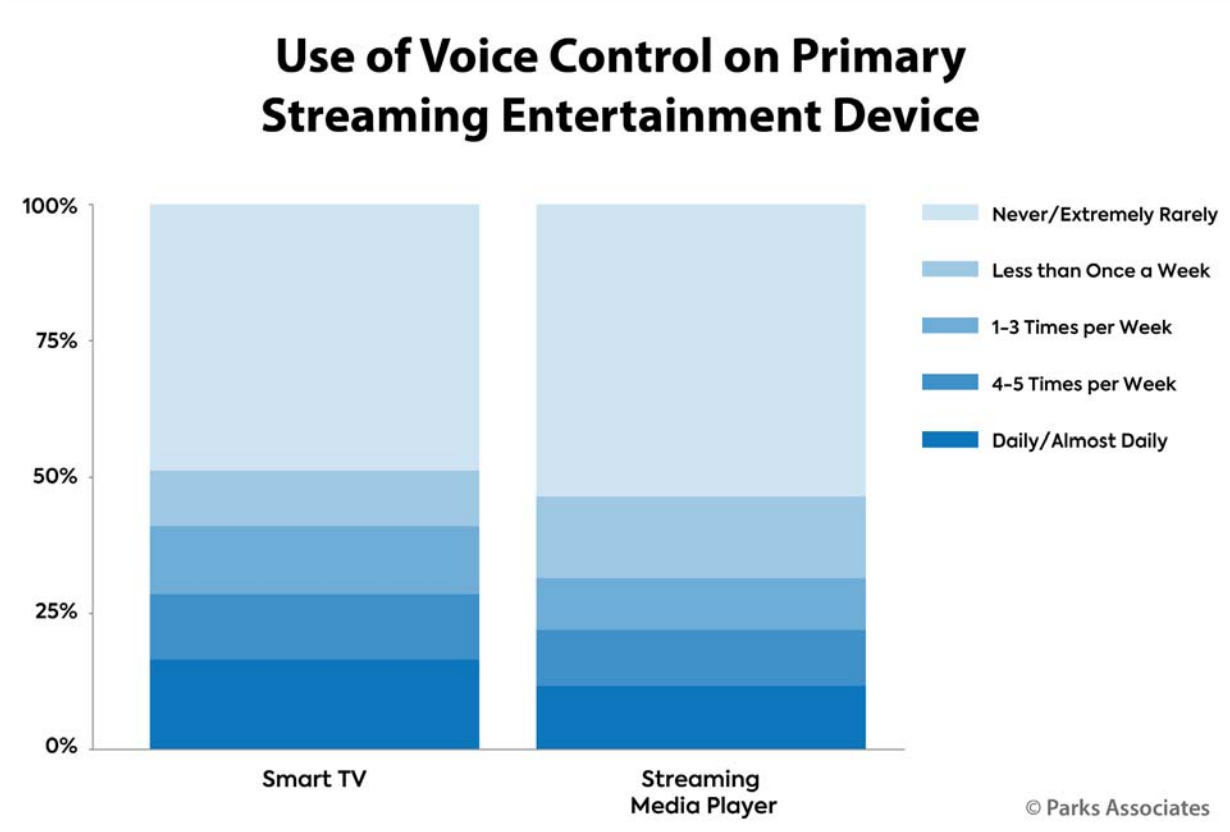
- 17% of smart TV owners use voice daily for control of the video experience.

Laptops

They are largely reliant upon two principal methods of input – the keyboard and the touchpad, or an attached mouse. Some models have augmented these methods with a touchscreen, cameras, fingerprint sensors, and microphones, but these devices overall remain principally bound to tactile methods of input.

- Laptops are in 77% of internet households.


Primary Control Method of Smart Home Devices
An app that controls a single type of product
An app that controls different types of smart products
My voice using a smart speaker/smart display with voice assistant
Physically touch controls on the device
My voice using voice-assistant on my smartphone
My voice using voice control that is built into the product
My security app on my computer, tablet, or smartphone
Panel or tablet from my security provider
My smart watch
Other
0%
35%
© Parks Associates


Use of Voice Control on Primary Streaming Entertainment Device
100%
75%
50%
25%
0%
Smart TV
Streaming Media Player
Never/Extremely Rarely
Less than Once a Week
1-3 Times per Week
4-5 Times per Week
Daily/Almost Daily
© Parks Associates

Smart Home Devices

- 38% of US Internet households own at least one smart home device, like a smart thermostat, smart door lock, video doorbell, or smart light bulb.

Dedicated apps are the primary control method among smart home device owners, but 28% say voice is their primary method of control via the device's native microphones and processing ability, or when issuing vocal commands to their smartphone, computer, tablet, or smart speaker (potentially with the help of a voice assistant). Voice is particularly popular with light bulbs and plugs where users may be operating devices while seated or with their hands full.

As noted, anything with a camera is a candidate for VBI, and some security systems are using embedded video to enhance the user experience, with facial recognition able to identify a household member versus a visitor versus an intruder. This could be an area in the smart home ecosystem where VBI could make early inroads, such as giving an identified head of household the ability to disarm or alert a system via specific eye movements or gaze.

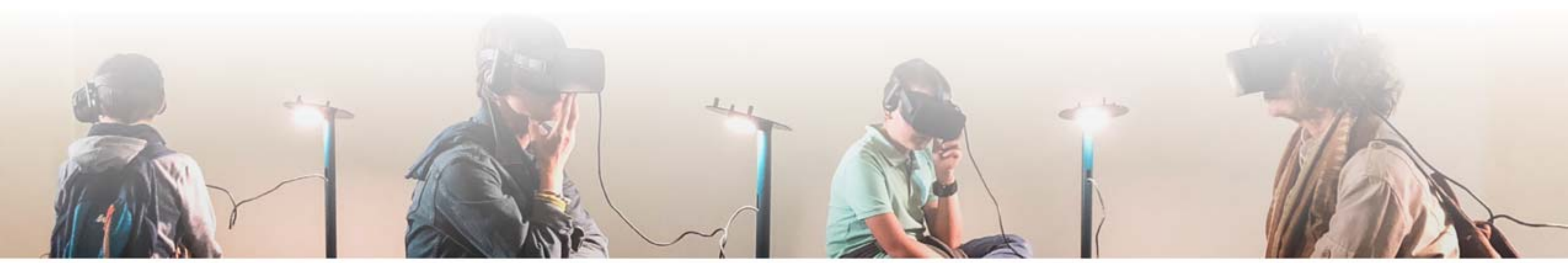
Virtual Reality (VR) Headsets

- 13% of US Internet households own a VR headset in 2022, double the adoption rate from 2019.

VR systems require multiple forms of input and positioning data to properly render a user in the virtual space. Accelerometers, gyroscopes, and magnetometers help to measure a person's rotational movement in space. Some systems also use special markers and depth-sensing cameras to obtain positional data. Accessories like hand controllers and gloves may provide additional input. Finally, eye tracking serves the purpose of aiming and navigating inside the VR headset display. Eye trackers measure both eye movement and the point in the distance at which the headset wearer's eye is focused (gaze point).

These devices are only a subset of the potential number of consumer technology devices that can benefit from vision-based interactions as a substitute for or augmentation of their existing touch-based input interactions. Their high level of adoption within society today is a hint of how broadly vision-based interaction capabilities can potentially make their way into consumers' everyday interactions with commonly accepted technology devices.

Vision-based interaction can be applied to these common devices and potentially augment and/or replace common interface modalities.



Vision-Based Interfaces

The ability of vision-based interfaces and interaction to augment and improve upon existing interface options is considerable.

Consumers are surrounded by an overwhelming array of devices, many requiring touch-based input. The use of vision enables an additional form of input that can augment or substitute for many of the same types of input possible via buttons and touchscreens. Any consumer device with a camera could be enhanced with VBI solutions. Additionally, vision-based input is not constrained by the same physical limitations / requirements as needing to touch something within arm's reach, and it has the potential to streamline interactions thanks to the speed and ease of utilizing a natural pointer – the human eye. The benefits extend beyond simple control, as VBI can serve as a gateway to AR (augmented reality) applications by delivering more relevant information to the user based on their focused gaze.

Many early use cases focus on users with disabilities that make the use of common input methods difficult or impossible. The CDC reports that 61 million adults in the US live with a disability, more than 13% with a mobility disability, nearly 6% who are deaf or have serious difficulty hearing, and more than 3% have a self-care disability with difficulty dressing or bathing.[1] Vision-based interfaces can create many new opportunities for this population to use and benefit from technologies previously unavailable to them.

Defining vision-based interfaces and related technologies

Vision-based interfaces provide hands-free control in place of or to supplement other forms of interaction and control input, such as tactile (touch) or vocal (voice) input. The principle of VBI relies upon the concept of CV.

Computer vision is a field within artificial intelligence (AI) that focuses on training computers to see, interpret, and understand the visual world based on observing and analyzing images or videos. It is used in essence to automate the tasks commonly performed by the human eye and brain, to process and react to what is "seen."

Vision-based interaction is reliant upon computer vision in that CV is being applied to the movements of the human eye. This use case of "eye tracking" entails measuring and interpreting what a user is looking at, their level of attention, and, in some cases, an intent or action.

This is accomplished with the use of one or more sensor inputs, commonly the combination of a camera and a near-infrared illuminator.

There are a variety of elements measured in the process of eye tracking:

- Gaze tracking and gaze point
- Head tracking and head pose
- Pupil detection and dilation level
- Eyelid opening and blink rate

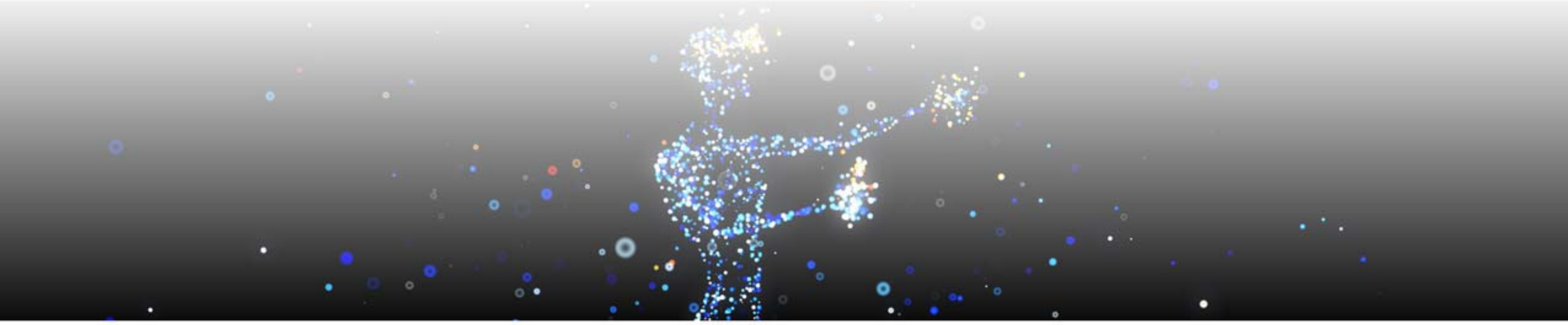
These examples are only some of the many individual elements monitored when attempting to assign meaning, context, intent, and action to human eye movement in relation to the surrounding world. Cameras and other sensors are always improving; as a result, eye tracking AI and algorithms are progressively "smarter" and better trained, processing power is always growing, and the intuitiveness of the interpretation of fused sensor data is always improving.

Common touch-based interactions applicable to vision-based interface

There are countless ways in which vision-based input can replace touch-based input today on mainstream devices:

- Keyboard and keypad inputs – users can visually select letters and digits to input without needing to be in physical proximity of a keyboard/keypad.
- Drag and drop – users can visually select an item on a screen by holding their gaze or blinking while focus is locked on an object on-screen, then use their gaze to drag / move it where desired.
- Dialog boxes – users presented with dialog boxes with simple yes/no questions or confirmation queries can quickly and efficiently navigate them.

These common interactions are only a select sample of the many different touch-based interactions that can be streamlined and improved with the addition of vision-based interaction.



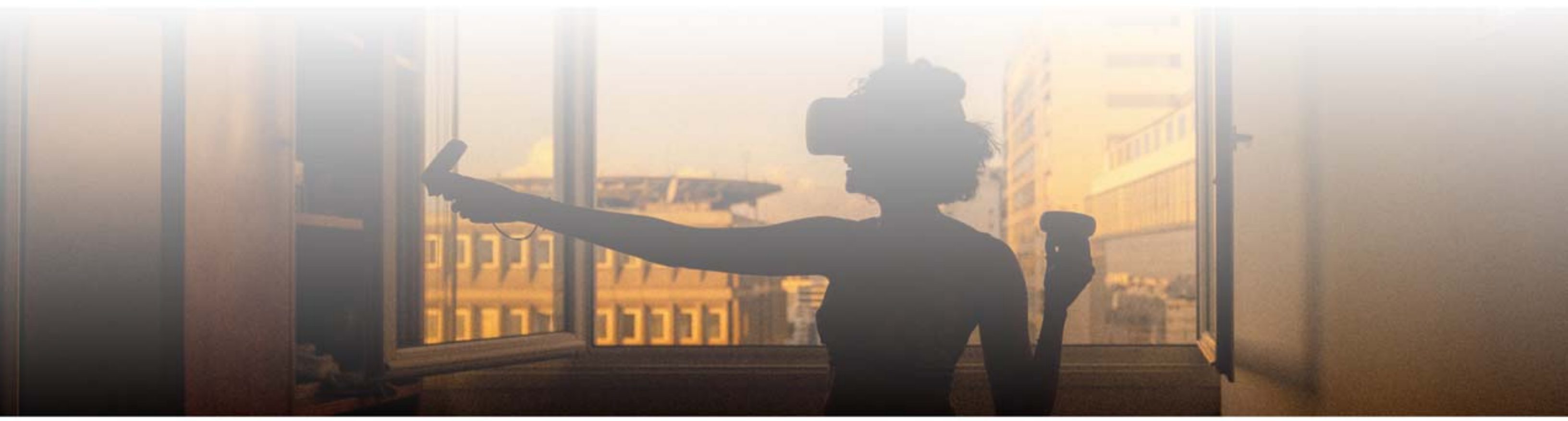
Vision-based interactions used with common consumer devices

Vision-based interaction and related technologies can be incorporated into many interactions that consumers have with technology and devices today.

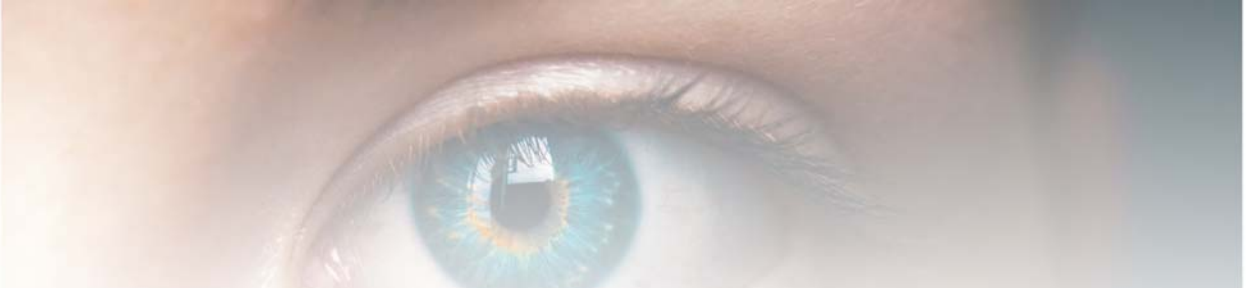
- **Connected home** – Today's smart homes are dominated by smartphones, smart speakers/displays, and voice control as inputs. Vision-based interfaces can allow lighting to be activated, devices to be turned on, and more, with just a gaze – and no voice or tactile input needed. VBI could even enable device pairing between interoperable devices, linking them with a gaze.
- **Virtual reality** – The gaming, social media, and entertainment markets are becoming ever more immersive, with tech giants building towards a "metaverse" concept. Vision-based interaction is becoming more common in the VR market, with eye tracking as most practical for foveated rendering and gaze-based control.
- **Smartphones** – Users can utilize vision-based inputs when touch and voice inputs are not possible or desired (e.g., wet/dirty hands or in a noisy environment). Beyond gaze-based typing, use cases could include answering or declining calls, opening a streaming music app and starting playback, taking a photo, recording video, and more.
- **Automobiles** – Vision-based interaction can be used in combination with advanced driving assistance systems (ADAS) to monitor the attention level of drivers: if their gaze is distracted away from the road, if they are falling asleep, and if an alarm/alert is necessary. VBI can also enable enhanced maps and navigation features.

Foveated rendering is based on the mechanics of human vision, where resolution is highest in the eye's fovea and lower in peripheral areas. Using this technique reduces the computational load by only rendering high-resolution images in gaze direction and providing lower resolution elsewhere. Foveated rendering is more realistic to human vision and saves processing power by only rendering in detail what the user is directly looking at. The solution provider can also collect data, via the VBI solution, on eye movements, length and subject of gaze, and the main objects/people of focus by the user. They can leverage that data to improve the experience and optimize content delivery by transmitting high-quality media to areas of focus and saving power through lower-quality rendering in the periphery.

Currently, gaze-based control is used particularly for aiming and target shooting in VR games.



adeia PARKS ASSOCIATES



Vision-based Interaction Advantages, Benefits, and Best Practices

The use of vision-based input and interaction as a control paradigm is a natural expansion of existing modalities of input. Vision-based interaction has the potential to drive several substantial benefits for both end users and industry, and this potential can be maximized with proper considerations during the conceptualization and implementation/execution stages.

Implications and benefits

Vision-based interaction brings a variety of benefits and advantages due to its ability to utilize a fast, natural input device – the human eye – to accomplish device control and input without having to commit the use of the user's hands, requiring a tactile input surface to be within reach, or requiring clear, audible voice input.

- **Speed and efficiency** – Vision-based actions can often be faster and more efficient than a touchscreen press or keypress on a keyboard. The ability to complete a task with merely one's gaze, if implemented correctly, can bring increased speed to common tasks or workflows.
- **Uninterrupted manual focus** – Use of the eyes as an additional means of input enables a heads-up mode of operation and allows the user's hands to remain in place for other tasks rather than disengaging to engage in tactile input, which can enhance safety and focus on important tasks.
- **Hygiene and epidemiological safety** – When direct contact with control surfaces is to be avoided, vision-based input can enable a higher level of safety for the immune-challenged or within research environments, pharmaceutical or medical device manufacturing, and other hygiene-sensitive use cases.
- **Accessibility and expanded choice of interaction** - Depending on the use case, vision-based input can serve as an augmentation to other interactions, a fallback/backup input method, and a way to expand input choices. This can be useful when a user has their hands occupied, the user has limited mobility or capability with their limbs, or any other circumstance where a single interaction modality is too limiting or too unaccommodating of user and/or situational variables.
- **Broad applicability** – In its most basic implementation, the principle of vision-based input can in essence be applied to nearly any device with a camera. Provided there is enough processing power to engage in computer vision processing, that device can potentially be operated without physical touch using vision-based input.
- **Authentication and personalization** – visual-based control technologies that can recognize unique irises may also be used for authentication, particularly for entertainment personalization. Usage of entertainment devices where multiple users are present differs from usage by a single user. By identifying the users present, features such as recommendations and other profile-oriented functions can be optimized for multiple users, enhancing the user experience.

This list is a limited set of advantages out of the many that are enabled using vision-based interfaces. The continual evolution of computer vision has enabled considerable advances in the quality and capabilities of vision-based interactions – but these capabilities must be teamed with intentional design and execution to deliver their full potential.

A poor user experience, especially early in deployment, could greatly inhibit VBI adoption and usage – it is critical that VBI is paired with a well-designed and intuitive user interface.

Best practices in the use of vision-based interfaces

The use of vision-based interaction can be useful in replacing or augmenting touch-based interactions, but its benefits can be maximized if intentional effort is made to follow known good practices.

- **Delivery of substantive benefit** – The use of vision-based interactions can only be an additive improvement or enhancement if it tangibly improves the quality and efficiency of the user experience. Two of the most fundamental aspects of the user experience that are commonly compared are speed and ease of use. Vision-based interaction should be implemented only if it improves the user experience in these and other fundamental dimensions for the use case. In other cases, vision-based interaction may be an even better fit than legacy methods of interaction for the particular use case, such as in sanitary environments.
- **Allowance for choice and contingency** – The variables inherent to computer vision, along with varying degrees of societal acceptance of the camera-based nature of vision-based interaction, mean that designing around only vision-based interaction for a particular use case is not a realistic approach. Allowing traditional input methods to also be used if desired provides both choice for users and a fallback input capability should environmental variables prove challenging to reliable vision-based input – necessary for mission-critical use cases.
- **Designing in fidelity, and utilization of sensor fusion where possible** – Vision-based interaction's ability can help build trust with the user, since the interface relies on accuracy and fidelity. Wherever possible, vision-based interfaces should leverage additional sensors if available in the operational environment (such as microphones, accelerometers, magnetometer/compass, etc.), for sensor fusion-based intelligence. This delivers two benefits; one is the ability to help cut down on false positives and ensure more accurate interpretation of eye-tracking-based inputs – increasing accuracy, fidelity, and user trust in the vision-based interaction. The second benefit is a potentially expanded command / input vocabulary that incorporates unique combinations of multiple input modalities, such as gaze with voice commands.

Rapid advances in processing power, camera sensors, and computer vision capabilities have substantially improved and are continuing to improve vision-based interaction capabilities. Realizing optimal benefit from these capabilities is only possible if the aforementioned practices and many more are observed during conceptualization and implementation. Well-considered design and execution are crucial to ensuring vision-based interaction delivers a net positive result for both users and stakeholders in real-world usage.

Next Generation of Device Interaction

The methods with which people interact with their devices are constantly evolving, and each change brings new benefits to everyday human-device interaction. Much of the technology for vision-based interactions is already in use today, and it is in active development and expansion, as multiple industries move to take the next step in the evolution of the user interface.

Multiple stakeholders within the technology industry, such as device manufacturers, UI and UX developers, camera sensor providers, computer vision software developers, and beyond – including touchscreen technology providers – are well-advised to actively take advantage of this trend towards incorporating vision-based technology into consumer device interactions.

There are also a number of challenges to consider, including costs to develop and implement into a product line, privacy concerns/concerns about new tech, and the hurdles of people learning and being comfortable with a new interface. Early deployments and trials will likely include a hybrid approach, where established modalities such as touch and voice are combined with vision-based interactions to deliver a new, optimized experience that provides the best combination of speed, ease, efficiency, and suitability to purpose. As noted, there are populations that will embrace the benefits and capabilities of this new method.

adeia

Adeia's vision-based technology capabilities currently enable key functions to enhance the entertainment experience at home:

- Users can pair two devices via Bluetooth through VBI
- User can then transfer content between the two devices, such as selecting a song from a smartphone to play on a connected soundbar, or displaying a map from a tablet onto the big screen of smart TV.

Ultimately, the goal is to deliver the best possible consumer device experience, and as technology continues to advance, the form that takes continues to evolve. Vision-based interfaces are a crucial element of this optimal vision, and their utility and usage are expected to grow over time.

[1] https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html#:~:text=61%20million%20adults%20in%20the,is%20highest%20in%20the%20South

adeia PARKS ASSOCIATES

About Parks Associates

PARKS ASSOCIATES

www.parksassociates.com
info@parksassociates.com
972.490.1113

Parks Associates, a woman-founded and certified business, is an internationally recognized market research and consulting company specializing in emerging consumer technology products and services. Founded in 1986, Parks Associates creates research capital for companies ranging from Fortune 500 to small start-ups through market reports, primary studies, consumer research, custom research, workshops, executive conferences, and annual service subscriptions.

The company's expertise includes new media, digital entertainment and gaming, home networks, internet and television services, digital health, mobile applications and services, consumer apps, advanced advertising, consumer electronics, energy management, and home control systems and security.

About Adeia

adeia™

Adeia invents, develops and licenses fundamental innovations that shape the way millions of people explore and experience entertainment and enhance billions of devices in an increasingly connected world. From TVs to smartphones, in almost any place, and across all types of entertainment experiences, from Pay-TV to OTT, Adeia's technologies allow users to manage content and connections in a way that is smart, immersive and personal. For more information, please visit **adeia.com**.

About the Authors

Jennifer Kent, Vice President, Research, **Parks Associates**

Jennifer manages the research department and Parks Associates' process for producing high-quality, relevant, and meaningful research. Jennifer also leads and advises on syndicated and custom research projects across all connected consumer verticals and guides questionnaire development for Parks Associates' extensive consumer analytics survey program. Jennifer is a certified focus group moderator, with training from the Burke Institute.

Jennifer earned her PhD in religion, politics, and society and an MA in church-state studies from Baylor University. She earned her BA in politics from the Catholic University of America in Washington, DC.

Tam Williams, Research Analyst, **Parks Associates**

Tam serves as a resource primarily to the Marketing and Research Teams, contributing to press releases, content marketing, event speaker logistics, marketing research, and many other projects.

With undergraduate and graduate degrees in business and theology, respectively, Tam's ongoing experience involves freelance writing and editing projects across multiple industries. University of America in Washington, DC.