Wearable Devices | mudra

A Wearable Devices Ltd.
White paper

INTERACTION BEYOND BOUNDARIES:
Elevating AR Glasses User Experience with Gesture Control and Neural Wristband

October 2024

© Wearable Devices Ltd. All Rights Reserved. Printed October 13, 2024

TABLE OF CONTENTS

PREFACE

Please feel free to reach out to Wearable Devices Ltd. at business@wearabledevices.co.il if you would like to provide feedback, corrections, engage in discussion, or share ideas regarding this publication. Please note that the use of the content in this white paper is subject to the limitations outlined in the agreement that readers accepted when downloading this document from the company's website.

This whitepaper consolidates three in-depth articles by our XR team. These reports were originally crafted as individual explorations, but by integrating them into a single document, we offer a comprehensive resource for readers. Some footnotes and images encourage readers to click links and view live animations and videos, which understandably are not available for print.

EXECUTIVE SUMMARY

The pace of technological advancement has always been dictated by user interfaces, and as the future shifts toward face-worn smart glasses, this whitepaper explores how to make these devices lightweight, comfortable, and user-friendly. Gesture control is ideal for face-worn devices, offering hands-free, intuitive interaction. The quality of input and display richness defines the user experience, making the right gesture selection crucial for natural interaction.

Two dominant approaches to gesture control are camera-based and wearable solutions, both facing challenges in accuracy and design. While camera-based systems are common in face-worn devices, we argue that IMU and SNC sensor fusion on a wearable wristband offers an equal or superior experience. Pointing and navigation are core to gesture control, and selecting the right gestures for screenless, monocular, and mixed reality displays is essential. With the Mudra Link, the world's first neural wristband, Wearable Devices Ltd. offers gesture control through wrist movement, finger gestures, and fingertip pressure. This solution provides familiar interactions like point, click, and drag while reducing the need for bulky hardware and extending functionality beyond the boundaries of camera-based gesture control solutions.

FORWARD-LOOKING STATEMENTS

This White Paper contains "forward-looking statements" within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, that are intended to be covered by the "safe harbor" created by those sections.

Forward-looking statements, which are based on certain assumptions and describe our future plans, strategies and expectations, can generally be identified by the use of forward-looking terms such as "believe," "expect", "may", "should", "could", "seek", "intend", "plan", "goal", "estimate", "anticipate" or other comparable terms. For example, we are using forward-looking statements when we discuss advantages and benefits of our Mudra Band and our Mudra technology, our belief that the Mudra Band is shaping and the landscape standard for the extended reality input and interaction and our mission to shape to shape the input landscape for extended reality. All statements other than statements of historical facts included in this White Paper regarding our strategies, prospects, financial condition, operations, costs, plans and objectives are forward-looking statements. Forward-looking statements are neither historical facts nor assurances of future performance. Instead, they are based only on our current beliefs, expectations and assumptions regarding the future of our business, future plans and strategies, projections, anticipated events and trends, the economy and other future conditions. Because forward-looking statements relate to the future, they are subject to inherent uncertainties, risks and changes in circumstances that are difficult to predict and many of which are outside of our control. Our actual results and financial condition may differ materially from those indicated in the forward-looking statements. Therefore, you should not rely on any of these forward-looking statements. Important factors that could cause our actual results and financial condition to differ materially from those indicated in the forward-looking statements include, among others, the following: the trading of our ordinary shares or warrants and the development of a liquid trading market; our ability to successfully market our products and services; the acceptance of our products and services by customers; our continued ability to pay operating costs and ability to meet demand for our products and services; the amount and nature of competition from other security and telecom products and services; the effects of changes in the cybersecurity and telecom markets; our ability to successfully develop new products and services; our success establishing and maintaining collaborative, strategic alliance agreements, licensing and supplier arrangements; our ability to comply with applicable regulations; and the other risks and uncertainties described in our annual report on Form 20-F for the year ended December 31, 2023, filed on March 15, 2024 and our other filings with the SEC. We undertake no obligation to publicly update any forward-looking statement, whether written or oral, that may be made from time to time, whether as a result of new information, future developments or otherwise.

PART 1: PROVIDING THE BEST INPUT EXPERIENCE TO INTERACT WITH SMART GLASSES

What's in a Name

Extended Reality. Spatial Computing. Ambient Computing. Augmented Reality. Virtual Reality. Mixed Reality. Metaverse. Did we miss something? Everybody is trying to name and claim dominance in the "next big thing" computer product category. Winning the customer's heart in this category will conquer their most desired user real estate: the face.

Juliet says it best: "What's in a name? That which we call a rose by any other name would smell just as sweet." Names are arbitrary labels and do not change the essence or true nature of something. The essence of a thing remains the same regardless of it's monkier.

Simply put, any face-worn device, from lean consumer stylish glasses all the way to a bulky large visor headset, allows the user to do two things: display various types of digital elements in the user's environment, and interact with these elements. **The richness of the digital overlays and the breadth of the interaction will determine the device's utility** - be it a practical tool for everyday use that remains on the face, or novelty hardware used for specific use-cases and is sent to collect dust on the shelf.

Some consumer level glasses only display short line text messages, with minimal block of the real-world environment, wherever the user is looking, and allow minimal or no interaction at all. Other high-level headsets display various screens, applications and widgets, enabling the user to pin certain display windows onto a specific location, and interact with multiple elements. With highly advanced passthrough cameras, users can adjust the amount of their physical surroundings they wish to view. With the plethora of AR and VR devices available in the market, it is important to first define these devices clearly before diving into a fuller discussion.

From AR to Spatial Computing

Just 12 years ago you could easily distinguish between face-worn device categories. Google Glass was the first consumer-level Augmented Reality (AR) Glasses, and Oculus DK1 were the first Virtual Reality (VR) headsets. There were several attempts at creating AR and VR consumer products by Vuzix and Sony, but none had achieved the same level of market readiness or mainstream appeal that Glass or Oculus did.

Mixed Reality (MR) became a thing in January 2015 when Microsoft announced HoloLens; Extended Reality (XR) became a thing just a year later, when Qualcomm launched the Snapdragon XR1 platform, dedicated for extended reality experiences and applications.

In 2021, Mark Zuckerberg rebranded Facebook as Meta to signify the company's expanded ambitions beyond social media and its commitment to developing the metaverse. In June 2023, Apple launched the first "Spatial computer" - the Apple Vision Pro.

In 2016 Passthrough was introduced in the HTC Vive. A front-facing camera enabled basic passthrough functionality - users can view real-world surroundings while wearing a device that blocks physical sight, overlaying digital content onto the real world. This innovation has blurred the previously distinct boundaries between AR, VR, and MR, merging elements of each and enhancing tasks like productivity, immersive gaming, and communication1.

AR. VR. MR. XR | explained.

Augmented Reality is the fusion of digital information elements with elements in the physical world. This heads-up display (HUD) type experiences overlay data without requiring users to look away from their usual viewpoints. These digital elements are basic, mostly in the form of icons or text, and present information that is not directly tied to the user's immediate physical surroundings. The display is designed to be visible without obstructing the user's normal field of vision, with a resolution of (480 to 640)×(360 to 850) pixels, on a monocular device which displays information to only one eye.

Virtual Reality involves creating an entirely digital environment that immerses users, completely replacing their physical surroundings. These immersive experiences typically use head-mounted displays (HMDs) to project the virtual world directly in front of the user's eyes. VR provides a fully immersive experience, often encompassing a 360-degree field of view. Users are surrounded by the virtual environment, which is designed to be as engaging and realistic as possible. The displays typically offer high resolutions, ranging from 1080×1200 pixels per eye, or even higher, on binocular displays - each eye has its own display, creating a stereoscopic 3D effect.

Mixed Reality is integration of digital information elements into the physical world, allowing for interaction between real and virtual objects. MR seamlessly blends digital content with the physical environment, enabling users to interact with both virtual and real-world objects simultaneously. It creates a dynamic experience where digital elements are contextually overlaid onto the real world, allowing for meaningful interaction between the two realms. Users may work simultaneously with various screens, applications and widgets, and allow the user to pin a certain element to a specific location. That element will remain where and when the user returns to the specific location. It usually uses high-resolution binocular displays, with 720 to 1080 pixels per eye.

Input Methods for Face-Worn devices

Input and interaction with face-worn devices can be accomplished through various methods, typically involving holding, touching, gesturing, or more recently, speaking to an interface. Alongside the traditional mouse and keyboard, you may find one or more of these input methods for smart glasses:

- **Touchpad -** mounted alongside the frame next to the head temple area, it allows users to navigate through the interface using simple gestures.
- **Wearable -** a ring, wristband or armband that uses wrist motion and finger movements to input commands.
- **Controller -** containing buttons and joysticks, touchpad controls, and even motion sensors which track hand and body movements.
- **Head tracking -** used to follow the user's head movements, adjusting and centering the display accordingly.
- **Gaze / eye tracking -** using eye movements or gaze direction for navigation.
- **Gesture recognition -** allowing users to interact with digital elements using natural hand or other bodily movements for direct manipulation.
- **Voice control -** allows users to interact with devices or interfaces by speaking commands
- **Hand tracking -** capturing the movement and position of a user's hands to enable gesture-based interaction.

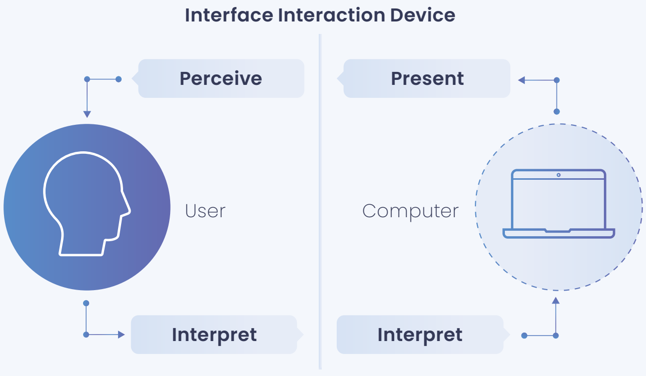
With the variety of input methods available for face-worn devices, such as gesture control, touchpad, and hand-held controllers, the overall user experience is shaped by more than just how users interact with the device. **The input method and the richness of the display determine the quality of core experiences with digital elements,** playing a crucial role in determining how immersive and effective the experience is. These factors work together to define how seamlessly users can engage with digital content, influencing their satisfaction and the effectiveness of the technology in various applications.

---

[1] What Provides the Best Experience to Interact with Smart Glasses? Wearable Devices' XR Team report, Mar 17, 2023

HCI, GUI, UX and UI

Human-Computer Interaction (HCI) is conveyed[2] through a user's input and computer output, by means of a user interface and an interaction device. The user forms an intent, expressed by selecting and executing an input action. The computer interprets the input command and presents the output result, which the user perceives to evaluate the outcome.



Human-Computer Interaction

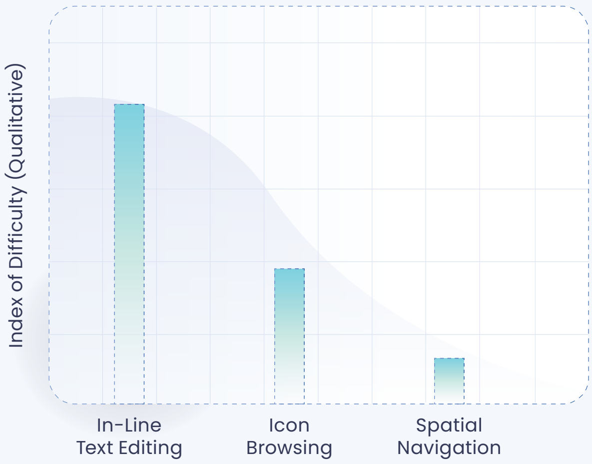
The goal is to make computer interaction feel Natural and Intuitive. **Natural** means you perform the input using a comfortable and relaxed body posture. You are at ease, and your body resides in a natural stance. **Intuitive** means you perform input using familiar and common methods. An intuitive interaction is one which binds the "same functionality" with the "same body movement" for any device.

In the context of User Interface (UI), a Human Interface Device (HID) is a type of hardware device that enables input and interaction between humans and computers. Traditionally, Input devices have been categorized as a Pointing Device or a Character Device. A pointing device, e.g. a computer mouse, is used to input position, motion, and pressure. A Character device, e.g. a computer keyboard, is used to input text.

Navigation and Pointing are the core functionalities of a Pointing Device. It comprises navigating in a 2-dimensional space, and manipulating - select or interact with - certain elements. Navigation and Pointing are the fundamental elements of pointing device human-computer interaction.

In the context of Fitts's law[3] in relation to a Graphical User Interface (GUI), spatial navigation requires the lowest level of cognitive load versus its peer groups of icon browsing and in-line text editing.

User experience (UX) in the realm of computer input encompasses the entirety of a user's interaction with a computer system. It incorporates a range of elements, including usability, accessibility, efficiency, satisfaction, and emotional response.



Fitt's Law and Index of Difficulty per interaction type

As mentioned above, there is a strong correlation between the richness of the display - hence the GUI - and the input method - the UI. Both have to adhere to Fitt's law hypothesis to provide a wonderful UX.

---

[2] Norman, D. A, (1984) Stages and levels in human-machine interaction. Int. J. Man-Machine Studies (1984), 21, 365-375

[3] Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement, Journal of Experimental Psychology, 47(6), 381-391

Pre-Digital to Modern Era Interfaces

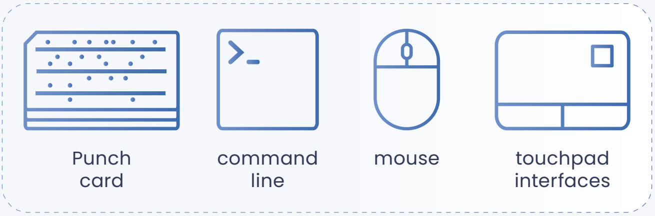
In the pre-digital era, the punch wheel was an early example of a mechanical device used to input data or instructions into machines. It featured a rotating wheel with holes or notches in specific patterns. Such patterns were read sequentially, allowing the machine to execute corresponding actions, a classic Turing Machine setup. An example of such devices are music boxes, where punch wheels control the sequence of notes played by aligning holes with a reading mechanism, producing programmed melodies.

The punch card emerged as a successor to the punch wheel, offering a more flexible and scalable method for data input and control in automated systems. It consisted of a stiff paper card with holes punched in specific positions to represent data or instructions, allowing for efficient storage, sorting, and processing.

The Jacquard loom, invented by Joseph Marie Jacquard in 1801, used punch cards as an input method to control the pattern of the weave. The punch cards encoded specific weaving instructions, allowing for the automatic production of intricate designs.

In the pre-digital era, expressing the intent was a physically laborious task with costly consequences in case of errors. The machine presented its interpretation as a form of a task such as looming fabric, and the user perceived the outcome and evaluated it through the finished product or result.

Keyboards became a standard interface in the 1970s and 1980s with the rise of personal computers and screens, allowing users to input text, execute commands, and navigate software. This era introduced the Command Line Interface (CLI), a text-based input method that allows users to interact with computer systems by typing commands into a terminal. It provided precise control over computing tasks, enabling users to perform file manipulation, system configuration, and automation through scripting.



The introduction of the Graphical User Interface (GUI) revolutionized computer input by enabling Direct Manipulation of visual elements on the screen, with the computer mouse playing an essential role. GUIs replaced text-based command input with interactive components like windows, icons, and menus. The computer mouse allows users to point, click, and drag these elements, making actions such as opening files, moving objects, and navigating applications intuitive and accessible.

Along with the computer mouse, common pointing device products are the directional pad and gaming controller. Newer input devices include the touchscreen, gesture recognition cameras and radars, IMU-bases wearables, and neural interfaces. As for character input, "speech-to-text" or voice assistants are now used to transcribe text or perform tasks based on verbal speech or command.

The pace of technological advancement has always been dictated by user interfaces.

The future of Computers is tilting increasingly toward face worn smart glasses, which require instant and relevant output interaction through a set of always-on sensors, based on the user's movement, aim and perspective – translating the user's intentions and inputs into actions.
Input interfaces for future devices will have to be re-invented as well.

Pointing devices: Left: A 1968 prototype of the first mouse (Getty Images);
1983 Nintendo Famicom controller with D_Pad (Nintendo);
(1977) Atari CX40 joystick (Atari)

ANALYZING APPLE VISION PRO, META ORION AR GLASSES GESTURE INPUT

Apple Vision Pro

One of the standout features of the Apple Vision Pro (AVP) is its use of hand gestures for interacting with VisionOS, delivering a smooth, intuitive, and seamless user experience. What makes this interaction feel so natural is its reliance on familiar gestures and its support for comfortable body postures, allowing users to engage with the system in ways that feel effortless.

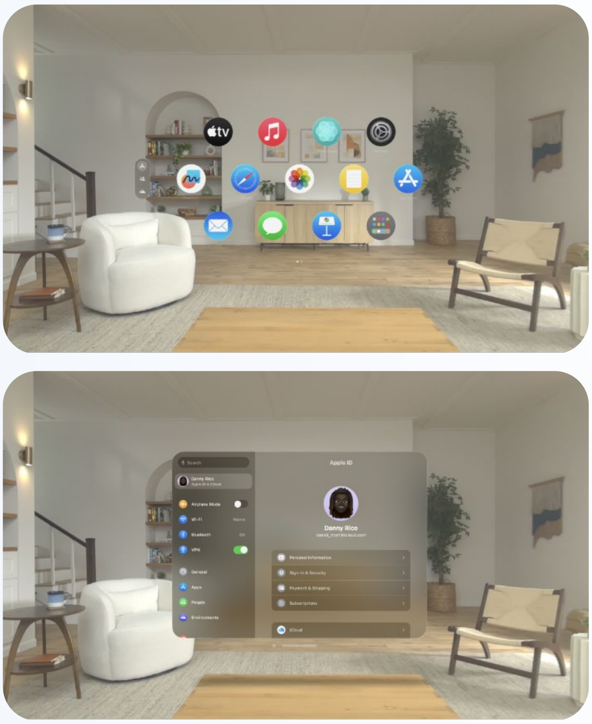
Interaction with the Apple Vision Pro involves the use of hands and eyes, along with the digital Crown and top button. The hands are used for Pointing, to input gestures such as tap, pinch and drag, pinch and flick, and virtual touch. These actions allow users to select an item, move items around, move or scroll content, and type on the virtual keyboard.

This is achieved by outward-facing cameras which track hand gestures, and with a large enough field of view and line-of-sight. The hand can rest comfortably on a desk or along the waist when making most gestures. The eyes are used for Navigating – the user gazes at an element, and it slightly changes its contrast or texture to hint that it is selectable.

Comparing the AVP input method with the HoloLens (HL1) first generation will immediately illuminate the lessons Apple implemented for its gesture recognition. In the HoloLens, the navigation was controlled by neck movements, with a fixed pointer in the middle of the display. The HL gesture camera's field of view was in the middle of the user's view - the user needed to place the palm of the hand in front of his nose, thus blocking the real-world view. Another AVP achievement is resolution and algorithms, which support very delicate and natural gestures, instead of the Bloom and Middle-Tap gestures used with HoloLens. Thus, Apple achieved both familiar and intuitive input gestures in comfortable natural body postures.

It's evident that the Apple Vision Pro inputs are commendable features that merit acknowledgement. However, there are aspects of the AVP interaction that warrant further exploration and refinement.

Using the eyes as an input interface presents challenges. While AVP's Home view offers large, easy-to-select icons with minimal eye movement, navigating the Settings menu requires precise focus and fine eye movements. This contrast underscores the importance of choosing the right input method based on task complexity, as discussed in GUI design and Fitt's Law. Eyes are not ideal as pointing mechanisms due to their role in sensing and cognitive functions, which limit precision. Additionally, prescription lenses are essential for accurate navigation, and recalibrating the eye and hand setup for each new user can be tedious, especially in shared usage.



Apple Vision Pro home view (right), settings screen (left).
Image source: Apple.com

Another major issue is comfort - weight and face fitment. The AVP comes with 2 types of Headbands, to provide cushioning, breathability, and stretch. It has approximately 30 shapes and sizes of face light-seal cushions to deliver a precise fit while blocking out stray light. Weighing-in at 650 grams and tethered to a 2-hour battery pack, it has been reported to cause users headaches, discomfort, and excessive sweating. According to the media, many Apple customers who returned their AVPs cited discomfort as a major issue. **It's 'wearable' only if users are willing to wear it.**

Why is it so bulky and heavy, with a relatively short battery life? That is partially because the complex gesture camera hardware and its algorithms require a lot of processing power and specific locations on the headset.

We believe that Apple hit the bullseye with the AVP's input User Experience – It employs the right gestures tied to the appropriate functions, offering a seamless, comfortable, and familiar input experience. **For most users, simple actions like point, click, and drag are almost all that's needed.**

Meta Orion AR Glasses

In Meta Connect 2024, Mark Zuckerberg unveiled Orion – 'a purposeful prototype' headset that combines augmented reality glasses, eye and hand tracking, and a gesture control wristband. Weighing around 100 grams (3.5 ounces), Meta's Orion AR glasses are designed to be lightweight and comfortable for extended use, while still packing advanced display and tracking technology for an immersive augmented reality experience.

Orion comes with three components: the glasses, the puck, and the wristband. The glasses use waveguide optics and micro-LED displays to project AR content. The puck functions as an external processing unit to handle computational tasks. The wristband tracks hand gestures for interaction with the AR system.

As published, navigation on the Orion AR glasses is achieved through gaze detection where the user's eyes focus on certain digital elements. The wristband is used for pointing.

Orion smart glasses, wristband and compute puck (Meta)

Meta neural wristband uses a Tap gesture to select and a Middle Tap gesture to invoke the glasses' main menu. Thumb Flicks - thumb pushes or pulls (as if tossing a coin) against the closed palm are used to scroll up or down. Double Thumb – a quick double thumb tap on the middle finger, is used to launch Meta AI.

A deep inspection into the nuts and bolts of the current Meta status may reveal it will take quite some time for their neural input wristband to hit the market.

The band form factor is that of a fitness tracker, and the design includes magnets and a clasp to keep a snug fit on the wrist. Several EMG sensors capture the EMG signals and translate them into gestures. The band also has haptic feedback, and processing is all done on the band itself, with a promised full day of battery life. The gestures can be done when the hand is completely resting alongside the waist.

Meta' EMG neural wristband (Meta)

---

[4] Introducing Orion, Our First True Augmented Reality Glasses

On an official blog post[4], Meta describes Orion as 'A Purposeful Product Prototype' which won't make its way into the hands of consumers rather they decided to focus on internal development first. Earlier this year, Mr. Zuckerberg stated that "we're actually kind of close to having something here that we're going to have in a product in the next few years"[5].

The gestures used for scrolling, which can be described as flicking or flexing fingers, are achieved by bending (or curling) the thumb finger toward the palm. Frequent or repetitive finger flexing may cause strain or discomfort over extended periods of use, especially in tasks requiring continuous fine motor control. The gesture could also interfere with other hand functions, making it less practical in everyday interactions.

In terms of gesture detection accuracy, some media reports mentioned that the gestures weren't yet perfect. Sometimes pinches were misinterpreted. That is probable, since flexing fingers may lead to unintended inputs, accidentally triggering a command.

While the haptic feedback provides a positive signal to reassure the user the gesture is recognized, it could suggest that the device may occasionally fail to detect gestures, leading to false negatives. This may indicate that the gesture recognition system might not be perfect, and the haptic feedback acts as a confirmation mechanism to reduce uncertainty for the user when a gesture doesn't register as expected.

As for the future of smart glasses[6], Mr. Zuckerberg says that the glasses are the next computing generation's phone, which will be used as an on-the-go computing platform. If Quest is the equivalent for your home screen TV, for most people the phone is probably the more important device in their life. However, in this next generation, glasses are probably going to be the more important and ubiquitous device.

Analyzing Meta's vision for interacting with smart glasses, we believe Meta has chosen the better approach and form-factor for inputting commands into face-worn devices. They are prioritizing face-worn, light and stylish form factor glasses. We believe that most users would prefer relocating input-related hardware to the wrist, eliminating the need for additional equipment on the face."

OPTIMAL UX AND UI FOR SMART GLASSES

Our exploration is finding the best experience to input commands into smart glasses. We started with a functional description of the face-worn product categories and the plethora of input methods used. We've established that the input method and the **richness of the display determine the quality of core experiences with digital elements.**

We then covered the HCI, GUI, UX and UI aspects for smart glasses, and defined what consists of a comfortable and familiar user experience. We learned that a pointing device's user experience can be tested by two parameters: navigation and pointing, and introduced the Index-of-Difficulty parameter which can be used to evaluate how natural and intuitive an input interface is. We also briefly covered the evolution of user interfaces, which led us to conclude that **the pace of technological advancement has always been dictated by user interfaces.**

We then moved to analyzing the interface input methods of the Apple Vision Pro and the Meta Orion AR glasses. We've pointed out that **Apple has mastered the optimal user experience input with its gestures, and Meta has chosen wisely with a wrist-worn form-factor gesture control interface.**

It is our belief that to become ubiquitous, Smart glasses need to be sleek, stylish, and lightweight. Using familiar and comfortable gestures for navigation and pointing to perform point, click and drag are the optimal input for rich display face-worn device.

Finally, shifting input-related hardware away from the face and making the device lighter and more comfortable to wear is essential for the widespread adoption of smart glasses.

Looks sell.

---

[5] Zuckerberg: Neural Wristband For AR/VR Input Will Ship "In The Next Few Years"

[6] Mark Zuckerberg Takes on Apple Fanboys, Tech Layoffs, Raising Cattle & More

PART 2: GESTURE CONTROL WEARABLE WITH POINTING DEVICE FUNCTIONALITY

Gesture Control - Origins and Evolution

Gesture Control technology allows users to interact with digital devices through hand, eye and body movements. It offers a natural and intuitive form of interaction. While nowadays the term is mostly related to the use of cameras or optical sensors, the first significant commercial application of gesture control using finger and hand movements was using a wearable[7].

Following the groundwork at Myron W. Krueger's artificial reality lab (1985), VIDEOPLACE, where users experimented interaction with computer-generated graphics using their body movements, the first significant application of gesture control technology was in VPL Research' development of the DataGlove in 1987.

DataGlove and Power Glove, late 1980's

The DataGlove was a wearable device that captured hand movements and finger positions, allowing users to interact with virtual environments through natural gestures. It employed fiber optic sensors to detect finger movements by measuring light transmission changes, a magnetic tracking system to determine hand position and orientation, microprocessors to process sensor data, and flexible circuitry to integrate these components while maintaining the glove's flexibility and user comfort.

The technology was licensed to Mattel which released the Power Glove in 1989 - a controller accessory for the Nintendo Entertainment System. It had traditional NES (Nintendo Entertainment System) joypad controller buttons on the forearm (directional pad and buttons), buttons labeled 0–9, and a program button. To input commands, the user pressed the program button and a numbered button. [Along with the controller, the player can perform various hand motions to control a character on-screen. It could detect roll, and uses sensors to detect four positions (2 bits) per finger for four fingers. Super Glove Ball, and Bad Street Brawler games were released with specific features for use with the Power Glove and included moves that can only be used with the glove.

While it sold nearly one million units and was a commercial success, because the controls for the glove were incredibly obtuse it became impractical for gaming. However, it was adopted by the emerging Virtual Reality community in the 1990's to interact with 3D worlds since it was cheaper than the DataGlove.

Power Glove, American model [Wikipedia]

In the following years, most of the significant research was conducted using cameras, in academic laboratories.

---

[7] Unlocking Gesture Control: The Rise of a Neural Input Wristband as the Next Generation's Pointing Device, Wearable Devices' XR Team report, May 27, 2024

Microsoft Kinect, 2010

Watch Video

In 2010 Microsoft launched the Kinect, a motion-sensing input device that enables users to control and interact with Xbox 360 games and applications through physical gestures and voice commands.

The Kinncet revolutionized gesture control by utilizing an RGB camera, depth sensors, and a multi-array microphone to track users' movements and gestures in three-dimensional space. This enabled full-body motion capture and voice recognition, allowing for interaction with games and applications through physical movements and voice commands without traditional controllers.

The Kinect key gestures were:

- **Wave:** Raise one hand and wave it side to side - used to start interactions or select items.
- **Push:** Extend your hand forward as if pressing a button - used to select or activate items.
- **Swipe:** Move your hand horizontally or vertically across your body - used to navigate menus or move between screens.
- **Raise Hand:** Lift one hand above your head and hold it - used to initiate interactions or bring up menus.
- **Grip/Release:** Close your hand into a fist to "grip" and open it to "release" - used to drag and drop objects.
- **Steering Wheel:** Hold your hands as if gripping a steering wheel and turn them - used to simulate steering in driving games.

The Kinect revolutionized human-computer interaction and gesture control by introducing motion-sensing capabilities to the mainstream. It enabled intuitive interactions using gestures and voice commands, significantly improving accessibility, and its versatile technology also found applications in virtual and augmented reality, expanding the scope of gesture control in various domains beyond entertainment.

Leap motion controller, 2013

Watch Video

In July 2013, Leap Motion launched its first product, the Leap Motion Controller, a groundbreaking device that allows users to control and interact with their computers using natural hand and finger movements. The Leap Motion Controller uses two infrared cameras and three LEDs to create an interactive 3D space, tracking the precise movements of the user's hands and fingers with incredible accuracy. It offered hand skeletal tracking data like the position of each bone of a finger or the orientation of the palm of the hand. It enabled touch-free interaction with a wide range of applications.

Gestures supported by the Leap Motion Controller:

- **Point:** Extend a finger to point and select items.
- **Pinch:** Pinch fingers together to grab and manipulate objects.
- **Swipe:** Move a hand or finger horizontally or vertically to navigate menus and screens.
- **Circle:** Move a finger in a circular motion to perform specific commands.
- **Grab:** Close a hand into a fist to "grab" objects and open it to release them.

The Leap Motion Controller revolutionized gesture control by providing high-precision tracking in a compact, affordable device, paving the way for new applications in various fields, including virtual reality, education, and digital art. It was compatible with HTC, Oculus and additional headsets and offered after-market gesture control functionality.

Myo armband, 2014

Watch Video

In 2014, the Myo Armband, a gesture control armband developed by Thalmic Labs, brought wearable gesture control technology back into the limelight. It used electromyography (EMG) sensors to detect muscle activity and motion sensors to track arm movements, allowing users to control digital devices through gestures. The Myo Armband's reintroduction of gesture control into the mainstream highlighted its potential across various applications, from gaming and presentations to drone piloting and virtual reality interactions, marking a significant advancement in wearable technology since the Nintendo Power Glove.

Gestures supported by the Myo Armband:

- **Double Tap:** tap your index on the thumb twice - used to select items
- **Wave Left:** Move your palm left - used for navigation or switching between items.
- **Wave Right:** Move your palm right - used for navigation or switching between items.
- **Spread:** Spread your fingers wide - used to pause or resume actions.
- **Fist:** Clench your fist - used to select items or perform actions like clicking.
- **Rotate, Pan:** wrist movements - used to adjust volume or scroll through lists.

The Myo armband gesture set

Microsoft HoloLens, 2016 **Watch Video** ▶

Microsoft' HoloLens 1, released in 2016, pioneered **built-in gesture control** in face-worn devices, allowing users to interact with holographic content through natural hand movements.
The original HoloLens used a few key gestures for interaction, primarily relying on simple hand movements to control the device. These gestures included:

- **Bloom:** This was the core system gesture, used to open the Start menu. The user would hold their hand up, palm open, then spread their fingers outward like a blooming flower.
- **Air Tap:** The Air Tap was used for selecting or clicking on items. Users would hold their hand up in front of the HoloLens, then tap their index finger and thumb together to simulate a click.
- **Tap and Hold:** Similar to the Air Tap, but users would hold their fingers together after the tap to allow dragging objects or interacting with more complex interface elements.

These gestures were the primary ways to interact with HoloLens 1's holographic interface, though they had some limitations in terms of precision and comfort compared to later developments.
HoloLens required users to position their hand in front of their nose to perform gestures, obstructing their real-world view because the gesture camera's field of view was at its center. Gesture control technology using internal cameras and sensors has since been adopted by devices like the Oculus Quest (late 2019), HTC Vive Focus 3 (2021), and culminated in the Apple Vision Pro (2023), hailed by many for its accurate, natural and intuitive gestures. These devices feature hand-tracking capabilities, enabling users to navigate interfaces, manipulate objects, and perform actions within virtual spaces using simple gestures for point, click, and drag functions.

Gesture control technologies and their boundaries

Gesture control technologies involve capturing movements, processing raw signal data by software to improve signal-to-noise ratio, and Machine Learning algorithms to interpret the processed signals to recognize and classify gestures.

For Vision-based gesture control, the hardware may include RGB Cameras that capture standard color images, Depth Cameras to measure the distance of objects, and Infrared Cameras to track movements in various lighting conditions by detecting heat signatures.

In addition, IMU (Inertial Measurement Units) sensors may be integrated to complement camera data with motion tracking. The software contains data processing firmware and signal processing libraries. The Algorithms may include Support Vector Machines (SVM) used for pattern recognition and classification tasks, Image Processing algorithms for Segmentation (Dividing) of images into meaningful segments to isolate hands and other relevant parts, Feature Extraction for Identifying key points and features in the images (e.g., fingertips, hand contours), and Tracking Algorithms (Kalman Filters): For predicting and smoothing the positions of moving objects for skeletal tracking reconstruction, and Machine Learning algorithms using Neural Networks to train models to recognize and classify gestures from visual data.

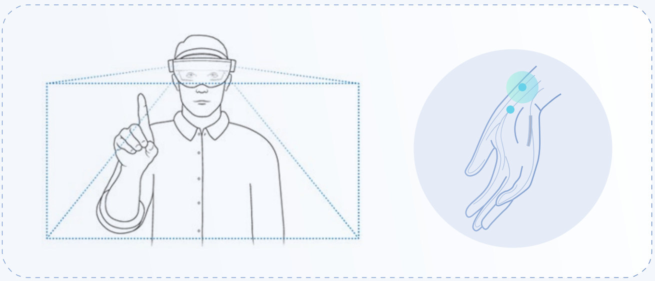
For Wearable-based gesture control, the general structure and types of technologies are quite similar, and the major distinguishable difference is at the hardware level. The wearable should be snugly fitted to the wrist to accurately detect bio-potential signals and wrist movement. The snug fit ensures minimal relative motion between the device and the skin or wrist, enhancing the precision of signal acquisition and motion tracking. The interfacing medium between the skin and the wearable device are the electrodes, which are used to detect electrical activity in the body.

Gesture control signal acquisition fundamentals

We now turn to the fundamentals of signal acquisition for gesture recognition. For vision-based technology these are known as **"Line of Sight"** (LOS) and **Field of View** (FOV), and for wearables the equivalents are **Electrode-Skin Contact Quality and Sensor Coverage Area.**

- **"Line of Sight"** refers to the requirement that the sensor or camera must have a clear, unobstructed view of the user's gesture to accurately detect and interpret it. If the view is blocked by an object the system may fail to recognize the gestures correctly.
- **"Field of View"** refers to the observable area that a sensor or camera can capture at any given moment, and it determines the complexity of the gestures that can be recognized. A wider FOV allows the system to capture more information, thus enhancing the accuracy and flexibility of the gesture or gestures.
- **Electrode-Skin Contact Quality** is crucial for accurate bio-potential signal detection, to ensure the electrodes can reliably measure bio-potential electrical activity without interference. Poor contact or obstruction will reduce the accuracy of gesture recognition.
- **Sensor Coverage Area** refers to the area on the skin where the electrodes can effectively detect signals. Proper placement and sufficient coverage are necessary to capture the full range of signal patterns and accurately classify the correlating gestures.

We can infer that the accuracy of gesture data collection is primarily influenced by the quality of the line of sight and electrode-skin contact, while the richness of gestures and the ability to classify them correctly are largely determined by the field of view and sensor coverage area.



HoloLens field of view (Microsoft) ; Sensor coverage area - illustrations

A good example of vision gesture recognition advancements in the past decade is comparing the HoloLens 1 gesture control with that of the Apple Vision Pro, as we've covered in Part 1 of this whitepaper, in the *Apple Vision Pro* paragraph.

This, of course, comes at a price. HoloLens 1 was priced at $3,499 back in 2016, which is the same price for the 2024 Apple Vision Pro. HL1 weighs 579 g (1.28 lb) and the AVP is around 650g (1.40 lb). The HL1 contained an internal rechargeable battery with average life rated at 2–3 hours of active use, or 2 weeks of standby time, whereas the AVP' external battery supports up to 2 hours of general use and up to 2.5 hours of video playback. Both devices have a "helmet" or ski-goggles form factor, with users complaining about the weight and front-loaded design.

Similarly, we may draw a comparison between the gestures of the Myo armband and the Mudra Link, while referring to the [a] electrode types and [b] coverage area.

The Myo Armband is worn on the forearm [b] and detects EMG signals [a]. It relied on broader arm and hand movements, such as making a fist or waving left and right, which required significant muscle activation and more noticeable gestures. It only featured discrete gestures.

In contrast, the Mudra Link uses SNC sensors picking neural signals [a] from the wrist [b], allows for much more subtle and natural finger movements, such as pinches and swipes, that are both delicate and intuitive. These finer gestures not only enhance comfort but also create a more seamless, familiar interaction, and it supports both discrete gestures such as Tap and continuous gestures such as Pinch and Drag. It can also measure fingertip pressure gradations.

To design a comfortable and familiar user experience with a new interface, it is essential to start by defining the desired interaction experience and then develop an interface that enables natural and intuitive interactions.

IMU and SNC: Tying Navigation and Pointing

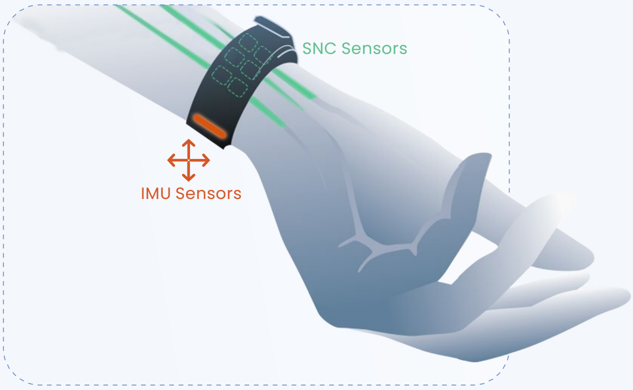
Gesturing can be performed either by vision-based input or by using a wearable for gesture control - voice may not serve the user well for pointing device functionalities but rather be more suitable for general actions such as "Hey Siri, launch the gaze-recognition calibration procedure".

Navigation and Pointing can be performed using a wearable by integrating two types of sensors: IMU for movement and an array of bio-potentials sensors for finger movements/innervation.

Inertial Measurement Unit (IMU) sensor is a device that typically combines multiple sensors, such as accelerometers, gyroscopes, and sometimes magnetometers, to measure and report information regarding the forces which act on the object, such as gravitational and geo-magnetic forces.

Surface Nerve Conductance (SNC) sensors are an array of electrodes that react to ions, via the process of ionic exchange and reacts to innervation picked up mostly by finger and hand usage patterns. The advantage of this sensor type is that it offers a non-invasive and convenient way to track physiological signals, such as fingertip pressure gradations.

A good combination of the two sensors allows for an accurate representation of both hand motion and fingertip pressure. Thus, the function of Navigation can be performed by the IMU sensor, and the function of Pointing will be handled by the SNC sensor. The IMU data could also be used in the Pointing algorithms.



IMU, SNC sensors locations on wrist

In our white paper **Designing a Neural Input Wristband for XR Experiences[7],** we've deeply elaborated on the importance of a proper wearable neural interface design for controlling an external device. We've argued that decoding the hand and finger intent movement requires sensor fusion, using sensors that can measure the movement of the fingers, the movement of the wrist, and the absolute true position of the hand is in relation to the body. Decoding the user's movement and intent depends on the correct sensors, as each sensor has advantages and disadvantages. Please refer to the SENSOR FUSION section in that publication for an introductory analysis on SNC and IMU sensors fusion.

---

[8] Designing a Neural Input Wristband for XR Experiences - a Wearable Devices white paper, Jun 5, 2023

In Depth Exploration of IMU and SNC Sensors for Gesture Control

Sensors such as IMU and SNC, integrated into a wristband form factor, form a crucial anatomical touchpoint. IMUs sense the motion of the wrist, as well as vibrations caused by fingers tapping together. SNC sensors can detect signals even when the wrist remains motionless, yet fingertip pressure is applied. When sampled in tandem, both sensor types provide arm movement, finger innervation signals, and noise. Each sensor has different properties. IMUs are well established MEMS sensors measure acceleration and angular velocity. SNC sensors are sensitive to the electrical activity of the nerves and muscles adjacent to the wrist.

To keep the discussion simple and coherent, we shall elaborate specifically on the “Tap” gesture – a soft tap of the index finger on the thumb - which is intuitive and is cognitively related to the action of choosing (or picking). A Tap movement creates a vibration, which can be sensed by the IMU. The vibration pattern is distinct; few other wrist movements can produce a similar pattern. However, individuals often have varying tap patterns, differing fingertip pressure levels, and distinct arm orientations. The tap gesture itself contains a large amount of variability.

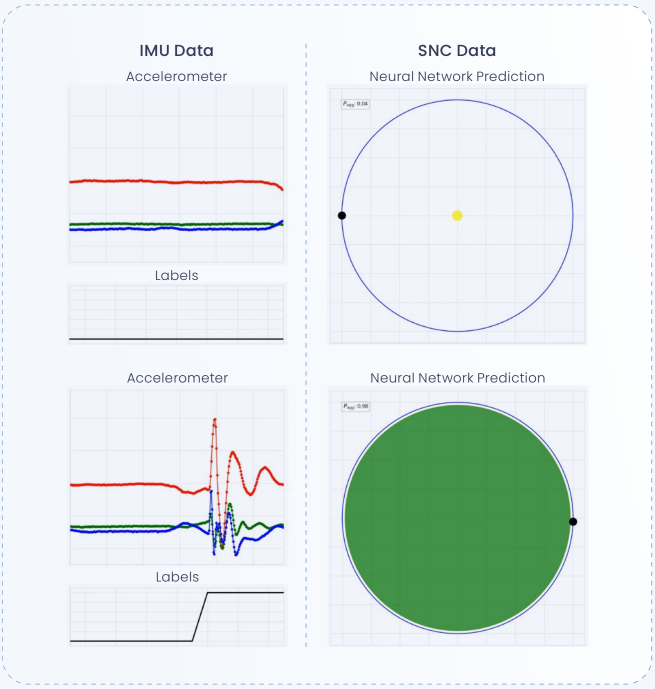
To collect data, we used a Mudra Development Kit wristband with high-speed IMU and SNC sensors mounted on it. Data was collected from multiple first-time users, while performing gestures of real usage scenarios – gestures such as tap, pinch-and-hold, and swipe. Additional recordings include “noisy” hand movements such as typing, drumming, walking.

To label the data, we’ve added an additional discrete sensor, mounted on the index and thumb fingertips, using a conductive material. When the index touches the thumb current passes through the fabric, thus providing both an external signal and physical force measurement.

Such a labelling mechanism will yield ‘1’ when the fingers are joined, i.e. a tap is performed, or ‘0’ otherwise. It provides an automated annotation mechanism for each of the above-mentioned gestures. This automatic method alleviates the need for laborious manual segmentation or implementing heuristic algorithms for labeling data. Thus, our approach leverages neural networks without the need for manual annotation.

IMU data - Tap Recognition

The neural network was trained using a large dataset collected from a diverse group of first-time users unfamiliar with the setup or gesture control scheme. As a result, the database includes a wide range of noisy inputs from various “worst-case” scenarios, such as different arm orientations, gestures, applied force, arm circumferences, and skin types.



IMU and SNC Data Visualization

To visualize the results, we’ve placed the IMU accelerometer data stream (left) alongside the neural network prediction (right), as can be seen on the following illustration. This example depicts a typical tap “tremor” classification (animation is available in the online version of the report).

On the top left, is an animation of accelerometer data, which is (part of) what the IMU “senses”, presented in red, green, and blue colors. On the bottom left, the labels denote the state of the button, presented using a black-color line. The right side displays an animation of the neural network inference. The compact, fully colored black circle tracks the state of the button. The inner circle denotes the probability of Taps using a real time aggregation neural network.

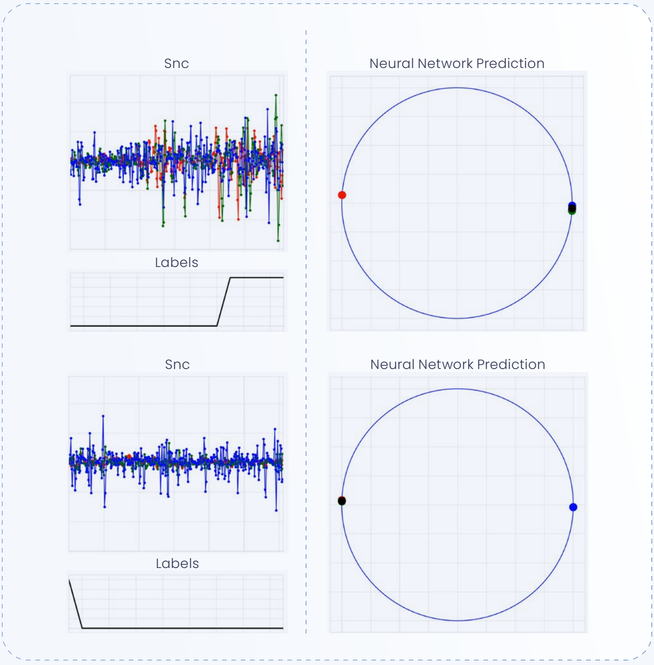
As sensor data represents a time series, decisions can be based on multiple successive windows (resembling several frames in the illustration). The window-wise accuracy measured on an internal database is 86%. When aggregating per-window results across multiple windows, with a second neural network, the accuracy rises to 96%.

A detailed discussion of this two-stage detector regarding Classifying taps using IMU[9] and Time series aggregation for real-time classification[10] is available.

SNC data - Fingertip Pressure Estimation

Our system measures biopotentials using an array of three electrode pairs. The signals presented show how pressing one's fingers together can be sensed, using a miniscule fraction of a camera sensor. Every subtle detail of innervation is picked up, along with noise and "noisy" movement.

The following illustration shows how applying fingertip pressure is visualized in the SNC signals array. On the top left side is display of the SNC sensors, colored in red, green and blue, respectively representing sensors SNC1, SNC2, SNC3. The data stream shows a drag-and-move gesture data acquisition sample. Such data is also a time series, yet it has different properties than the Tap gesture: It is indefinitely longer time than a momentary tap, and electrode disconnections, motion, and friction artifacts, may occur during this time. At the bottom left, the labeling is '1' when the user applies fingertip pressure. In general, higher amplitude SNC signals (at certain frequencies) are indicative of innervation.



SNC continuous fingertip pressure visualization

Combined movement and pressure is generally noisy: the electrodes may dis-attach from the wrist skin surface thereby introducing motion or friction artifacts, as well as differences in user behavior, the amount of pressure applied, and user-physiology skin impedance variances. Not all electrodes are in snug fit (constant contact) with the skin, which significantly increases the difficulty of classification.

Each electrode is mapped to a corresponding state on the large circle on the right, colored accordingly. Using a majority vote overcomes disconnections in the SNC sensor array, by ignoring noisy electrode inference. Overcoming disconnections in SNC sensors is deeply elaborated in this source[11].

Training a neural network on each sensor in this array is done separately, which yields a per-window accuracy of 83%. The results are aggregated from all the sensors together and the use of majority vote across the sensors yields an improvement of 89%. Results are also aggregated in time, with such aggregation requiring more advanced memory mechanisms. When aggregating in time using custom recurrent cells, we observe an accuracy of 94%.

---

[9] Afanaseva, S. Wearable Sensors for Effortless Interaction

[10] Afanaseva, S. Tap recognition task: Second Stage, Aggregation Problem

[11] Afanaseva, S. Tap recognition task: Second Stage, Aggregation Problem

Extending IMU Data beyond field of view boundaries

Comfortable interaction is crucial for input, and while many envision "Minority Report" style gesture control, holding hands mid-air is far from natural. *In fact, during filming, Tom Cruise needed riggers to suspend his wrists like a marionette because his arms became so fatigued from the opening scene that he could no longer lift them (aka "gorilla arm syndrome").*

Tracking the user hand movement is challenging to the camera due to the line-of-sight limitation. Placing an outward-facing camera array is not a solution for lightweight face-worn devices.

Early VR systems used "outside-in" tracking with fixed cameras positioned in the surrounding environment (e.g. in a playroom) which does not constrain the user to FOV limitation. The obvious disadvantage is that such a system is completely stationary after installation.

XR gaming controllers, hand-held input interfaces, use "inside-out" tracking; An IMU can perform such tracking by estimating orientation and positioning. However, the quality of such tracking is limited, based on the noise properties of the IMU itself and the estimation algorithms used.

The key advantage of a wearable interface is that it is completely free of such restraints, such as stationarity, FOV and line-of-sight limitations. Yet, it provides the same user experience.

IMU data derived from the wrist can be used for input motion and navigation, like an Air mouse, which gives a cursor control by moving a hand-held device in the air. It uses motion sensors, without the need for a flat surface like traditional mice. With a wearable interface, we move from hand-held to touchless interaction. Thus, using a built-in IMU on a wearable wristband provides navigation beyond the boundaries of camera-based line of sight.

Gesture Control Wearable as the Next Gen. Pointing Device

A neural-based wearable wristband can provide users with a pointing device's functionality of point, click and drag while maintaining high accuracy, user comfortability, and glasses like face-worn device form factor. **Form follows function.**

In the past 35 years gesture control technology has been alternating between wearables, external sensors, and built-in camera arrays. It has moved from large mid-air indirect arm and hand gestures to fine, subtle direct tap, pinch and drag using the index and thumb.

For camera-based gesture control, line of sight and field of view determine how comfortable the body posture and familiar the gesture can be. We've correlated these factors with electrode-skin contact quality and sensor coverage area for neural interfaces.

IMU and SNC signals collected from the wrist can provide the point, click and drag functionality using simple gestures like wrist movements, taps, and applying fingertip pressure while moving the wrist. A neural wearable can provide interactions beyond line-of-sight limitations.

IMU can be used specifically for navigation in "Air-mouse" mode, providing the key advantage of not being limited by the gesture recognition camera array line of sight and field of view.

A wearable gesture control is the perfect companion for slick, lightweight and all-day smart glasses.

Design drives demand.

PART 3: INTEGRATING WEARABLE AND CAMERA FOR GESTURE CONTROL

Tying Display Richness and Pointing Device Type

As discussed in part 1, GUI design for face-worn devices is quite versatile, and is mostly dependent on the display size and resolution, which determine the device form-factor and dimensions. To illustrate this, let's briefly explore input functionality for three types of devices: a screenless glasses device, a monocular AR heads-up display, and a MR device.

Please visit the online version of this report to view live animations of the gesture types[12].

Meta Ray-Ban

These are glasses that are designed to provide a seamless experience for capturing moments, staying connected, and enjoying media. They feature dual 12MP ultra-wide video recording cameras, integrated open-ear speakers for audio playback, and multiple microphones for voice capture and calls. The Input method for the device includes a touchpad located on the right temple, utilizing swiping and tapping gestures.

The main navigation functions are:

- **Single Tap -** to Play/Pause Audio or Answer/End Call
- **Double Tap -** to Skip to the next audio track.
- **Triple Tap -** to go back to the previous audio track
- **Swipe Forward -** to increase the volume
- **Swipe Backward -** to decrease the volume.

EverySight Raptor

These are a set of augmented reality (AR) smart glasses designed specifically for cyclists, featuring a monocular heads-up display (HUD) of 872x500 pixels. The input method for the device is a touchpad in the forward portion of the arm at the right temple, controlled using swiping and tapping.

The main navigation functions are:

- **Swipe forwards or backwards -** to rotate the carousel, using one finger in a long continuous swipe motion.
- **Tap -** to select an item. The item selected is always the one that is centered.
- **Swipe down -** to go back one screen.
- **Tap and Hold -** to show the list of running apps and settings/adjustments.
- **Double Tap** - to activate the camera in.

Microsoft HoloLens 2

A set of mixed reality (MR) smart glasses designed specifically for professional use, featuring a dual 2K 3D holographic resolution per eye. The input method for the device includes advanced hand tracking for intuitive interaction and eye tracking control. Hand gestures are used to interact with holograms:

- **Air Tap -** to select items
- **Bloom -** to open the Start menu
- **Tap and Hold -** to dragging and manipulating objects
- **Hand raise -** to target and interact with distant objects
- **Scroll -** to scroll through lists and content
- **Pinch to Zoom -** to zoom in/out content

The Meta Ray-Ban glasses have no screen, the Raptor has a minimal display screen, and the HoloLens 2 has a very rich display. There are many similar products in the market that use similar input methods for the same display type.

Defining Device Display Categories

As the display becomes richer, the input interactions naturally grow more complex and sophisticated, demonstrating a clear and strong correlation between GUI and HCI.

---

[12] Interaction Beyond Boundaries - Integrating Wearable Neural Interfaces and Gesture Recognition Cameras for Enhanced User Experience, Wearable Devices' XR Team report, May 27, 2024

We can classify the display types of these three product categories in the following way:

Screenless glasses, e.g.: Meta Ray-Ban glasses, Snap Spectacles 3, Bose Frames, Amazon Echo Frames.

Monocular devices, e.g. EverySight Raptor, Google Glass Edition 2, Vuzix M400, Epson Moverio BT-40S, RealWear HMT-1.

Mixed Reality headsets, e.g. Microsoft HoloLens 2, Apple Vision Pro, Magic Leap 2, ThinkReality A3.


Minimal Display
Monocular Display
Rich Display

The green area exemplifies the richness of display per product category (Wired)

The Meta Ray Ban glasses GUI uses audio and action-results cues for the user to receive feedback. Raptor's GUI features an Icon-Based Navigation system with a grid layout of icons to navigate through. HoloLens2 uses a full Graphical User Interface, i.e. 2D/3D spatial navigation.

What are the classical types of pointing devices suitable for each product category? We may consider the following input scheme:

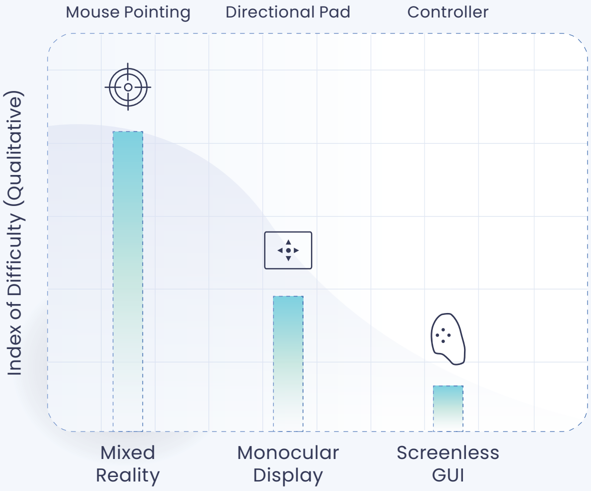
- Screenless glasses can be controlled using Controller functionality. Using an analog stick for swiping and buttons for selection and skipping.
- Monocular devices can be controlled using a simple **Directional Pad** functionality. Using arrows or a pad for navigation and button or tap for selections.
- Mixed Reality headsets can be controlled using a **Mouse Pointing** functionality. Moving a pointer for navigation, buttons for selection and both for dragging.

As discussed in part 1, input interaction with a GUI, in the context of Fitt's law, can be more easily understood using the metaphor of driving a car.

Inputting commands into a screenless GUI is like driving on a highway - only a slight nudge of the steering wheel is required to get back on course. In the same way, simple finger movements while the wrist isn't moving can be used when activating a controller to interact with a screenless GUI.

Inputting commands on a monocular device requires a bit more user attention - browsing through icons, which is analogous to driving in urban areas. Navigation is through large icons when the selectable element is highlighted or centered on the display.

Driving inside a crowded garage requires a good degree of focus, which is relevant when interacting with small digital elements or inline text editing on Mixed Reality headsets.



Qualitative Index of Difficulty, Display Category and Input Type

The Preferred Gesture Types per Device Display Richness

For screenless GUI devices, which can use minimal input schemes, one can **use discrete finger movements** to input commands. A tap, a double-tap, and index or thumb finger movements with the wrist orientation up or down can provide 8 input gestures, which is well beyond the scope for navigation and selection, and aligns with Miller's number of objects an average human can hold in short-term memory[13]. The benefits of using such

---

[12] Miller, G. A., "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review. 63 (2): 81–97(1956)

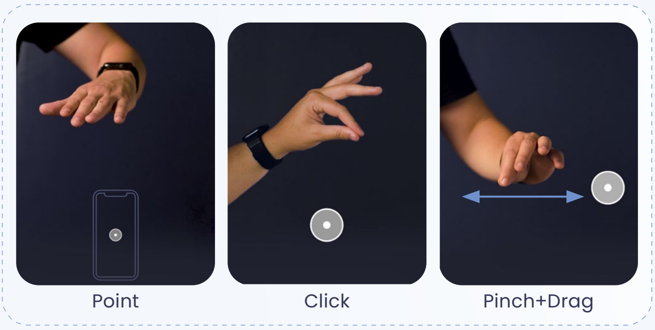
gestures is that they do not require high attention nor display feedback.

Discrete gestures input for minimal or screenless devices

For monocular displays, **using flicks - hand swipes and scrolls for navigation** - and tap/ double tap gestures to input commands. The subtle visual feedback of the selectable highlighted icon provides a simple and tactile gesture input design, which offers clear, distinct directional gestures that minimize the likelihood of accidental presses.

Flicks and Taps for monocular or heads-up displays

Mixed Reality and rich displays (e.g. VR/XR/ Spatial, Smart TV) **pointer and gestures offer a comfortable and familiar interface** and a great user experience. Controlling a cursor using small wrist movements, tapping on objects, and drag element manipulation using a pointing combination of wrist movement and fingertip pressure is probably the most desirable control method. It has been adopted by the most advanced MR and Spatial devices in the Market.



Pointing device functionality gestures for Rich displays

Now that we've matched the optimal pointing device functionality with each product category, let's examine how to provide an enhanced gesture control experience.

Enhancing User Experience Through the Blend of Wearabl and Camera Gesture Control

In June 2024 wer'e demonstrated Mudra technology on Lenovo's ThinkReality A3 headset, at the Augmented World Expo (AWE).

The Lenovo ThinkReality A3 smart glasses are advanced augmented reality (AR) eyewear designed for professional use. They feature high-resolution stereoscopic 1080p displays, providing a virtual screen experience equivalent to viewing multiple monitors. The glasses are powered by the Qualcomm Snapdragon XR1 platform and offer integrated 8MP RGB cameras for video recording and streaming, along with dual fish-eye cameras for room-scale tracking. The ThinkReality A3 is lightweight and can be connected to a PC or select Motorola smartphones via USB-C, making them versatile for use in various business applications, including virtual monitors, remote assistance, and 3D visualization.

The A3 smart glasses provide a variety of input methods to improve user interaction, including voice commands, head movement tracking, and gesture control via the camera. We've demonstrated three input modalities which enhance the user experience and show how the use of camera and wearable gesture control are mutually compatible.

1. **Blended modality -** Mudra input works alongside the A3 head tracking. The user navigates to areas of interest with the A3 head movements, and pointing is achieved using Mudra tap or pinch and drag gestures. This empowers comfortable body postures for input without raising the arms in mid-air.
2. **Extended Input -** we've used Mudra input to enhance interaction with the glasses beyond the field of view boundaries. While inside the FOV the A3 gesture control is used, and outside the FOV a laser pointer is controlled by Mudra. This benefits the user with more streamlined interaction without the constant need to move the head.
3. **Air stylus -** Mudra technology has the unique ability to measure fingertip pressure gradations.

   We've created a digital art experience which lets users draw in mid-air simply by moving the hand and control the width of the line by applying various gradations of fingertip pressure. This method offers enhanced freedom and flexibility, enabling more dynamic gestures by eliminating the need for a stylus, reducing hand strain, and providing precise sensitivity for accurate, natural-feeling drawing.

Gaze/eye tracking integrated with neural wristband for smart glasses interaction

Wearable Gesture Control Enhances the Capabilities of Face-Worn Devices

On average, a pair of sight glasses typically weighs between 20 to 50 grams (0.7 to 1.8 ounces), while a pair of sunglasses generally weighs between 25 to 50 grams (0.9 to 1.8 ounces). Snap Spectacles weigh 56.5 grams, and Meta Ray-Ban glasses range around 49 grams. Devices are priced at around the $299 mark.

For AR glasses to achieve mass market adoption, they must be lightweight, comfortable with enough 'juice' for all-day wear and continuous use.

We believe that a neural input wristband can offer the same accuracy and experience as that of the Apple Vision Pro, while massively reducing the glass' weight, form factor and price, and a battery that can last longer.

Dare to go beyond.

CONCLUSION

This whitepaper outlined a path for making face-worn devices lightweight, comfortable, and user-friendly by shifting input control to a gesture-based wristband. Face-worn computers require seamless input and output, with interactions driven by the richness of their displays.

Our user interaction analysis of Apple Vision Pro and Meta Orion AR glasses suggests that familiar tap and drag gestures, combined with wrist movements, offer the optimal user experience, and can replace the reliance on eye/gaze tracking for navigation.

As gesture control has evolved from large body and arm movements to fine, subtle gestures, we found that different face-worn devices benefit from tailored input methods. Discrete taps and wrist orientations suit screenless glasses like Meta Ray-Ban, while flicks and swipes are ideal for monocular products like Vuzix M400. For mixed reality devices like Apple Vision Pro, point, click, and drag gestures provide excellent control.

The market is clearly ready to embrace the fusion of gaze tracking for navigation and gestures for pointing interactions. Users are increasingly accustomed to the seamless integration of these natural human behaviors, as gazing and hand movements are rapidly becoming the new standard for immersive experiences. The Mudra wearable input solution enhances the user experience by providing multiple integration modalities for navigation, pointing, and interaction, alongside camera and voice input solutions.

The Mudra Link neural interface wristband offers all three input modalities—mouse, directional pad, and controller—delivering high accuracy and all-day comfort. Its multiple discrete gestures, fingertip pressure, and wrist movements enable a versatile input solution

By relocating input functions to the wrist, the form and weight of face-worn devices are reduced, improving wearability and adoption. We invite you to explore how neural wearable technology can enhance your products and services - **dare to go beyond boundaries!**

#BeyondBoundaries

MUDRA LINK – UNIVERSAL GESTURE CONTROL NEURAL WRISTBAND

Mudra Link is a revolutionary neural interface wristband that allows you to control your computers and devices using nothing more than simple hand gestures. Mudra Link is not just another wearable; it's a groundbreaking extension of your body, translating subtle movements into powerful commands.

The Mudra Band's Air-Touch functionality offers high accuracy navigation and pointing through multiple discrete and continuous finger micro-gestures. Combined with wrist movements, it offers multiple tap, pinch, drag and flick gestures sets, to cater all input modalities face-worn devices.

The Mudra Link technology and functionality has a proven market track record as it has been adopted, tested and used by thousands of Apple enthusiasts who purchased the Mudra Band, our flagship product which share same core technologies with the Mudra Link.

What Makes Mudra Link Stand Out?

- **Universal Gesture Control -** Whether it's skipping a song on your phone, answering calls, or controlling a presentation, all it takes is a simple gesture. Wearable, Touchless, Hands-free.
- **Neural, Natural, Intuitive -** an extension that bridges the physical and digital realms using familiar and comfortable gestures that respond instinctively to your actions.
- **Mudra Link Mapper -** Chart your own path by customizing gestures to specific commands for your devices. Explore new horizons in human-computer interaction with comfort and ease.
- **Sleek Ergonomic Design -** ergonomic design ensures comfortable wear with a sleek, modern aesthetic look that fits naturally both for professional and casual settings.

Mudra Link embodies this future, allowing one to control devices through simple gestures, seamlessly integrating technology as an extension of ourselves. This is the dawn of the human era of technology—where devices respond to our natural gestures effortlessly. We invite you to embrace the future of human-tech interaction with Mudra Link today.

Visit: https://mudra-band.com/pages/mudra-link-main

ABOUT WEARABLE DEVICES

Wearable Devices Ltd. (the "Company") is a publicly traded technology growth company specializing in artificial intelligence ("AI")-powered touchless sensing wearables for the B2C and B2B markets. The Company's flagship product, the Mudra Band for Apple Watch, integrates innovative AI-based technology and algorithms into a functional, stylish wristband that utilizes proprietary sensors to identify subtle finger and wrist movements allowing the user to "touchlessly" interact with connected devices. Recently the company has launched the Mudra Link, a neural interface gesture control wristband which to universally controls various computers and devices of various makes and operating systems.

The Company also markets a B2B product, the Mudra Development Kit, which utilizes the same technology and functions, and is available to businesses on a licensing basis. Wearable Devices is committed to creating disruptive, industry leading technology that leverages AI and proprietary algorithms, software, and hardware to set the input standard for the extended reality experiences and face-worn computers, which are rapidly expanding in the consumer markets.

The Company's ordinary shares and warrants trade on the Nasdaq Capital Market under the symbols "WLDS" and "WLDSW", respectively.

Headquartered in Israel, the company was founded in March 2014 by Asher Dahan, Guy Wagner, and Leeor Langer, with renowned expertise in analog sensors, signal processing, deep learning and machine learning algorithms, software, and human-computer interaction.

All content and materials included in this white paper are protected by copyrights and are proprietary to WEARABLE DEVICES LTD. Except as expressly authorized by WEARABLE DEVICES LTD., you agree not to distribute, publish, copy, modify, reproduce, transfer, publicly display, adapt, edit, or create derivative works from such content or materials. Any rights that are not expressly granted to you are hereby reserved by WEARABLE DEVICES LTD.

Notwithstanding the above, you may print or download one copy of the white paper on any single computer for your personal, non-commercial use, provided you keep intact all copyright and other proprietary notices.

The content, information, opinions, and viewpoints contained in this white paper are those of the authors of this white paper. Such information is provided "AS IS" and is believed to be accurate at the time of its publishing. WEARABLE DEVICES LTD. MAKES NO WARRANTY, EXPRESSED OR IMPLIED, as to the completeness or accuracy of the content contained in this white paper. The reader uses this white paper at his or her own risk, and WEARABLE DEVICES LTD. shall not be responsible for any errors, omissions, or inaccuracies in this white paper, whether arising through negligence, oversight, or otherwise. Reliance on any information appearing on this white paper is strictly at your own risk.

This white paper does not take into account, nor does it provide any legal or investment advice or any opinion regarding the specific investment objectives or financial situation of any person.