Human Computer Interaction

Fundamentals and Practice [SWE - 431]

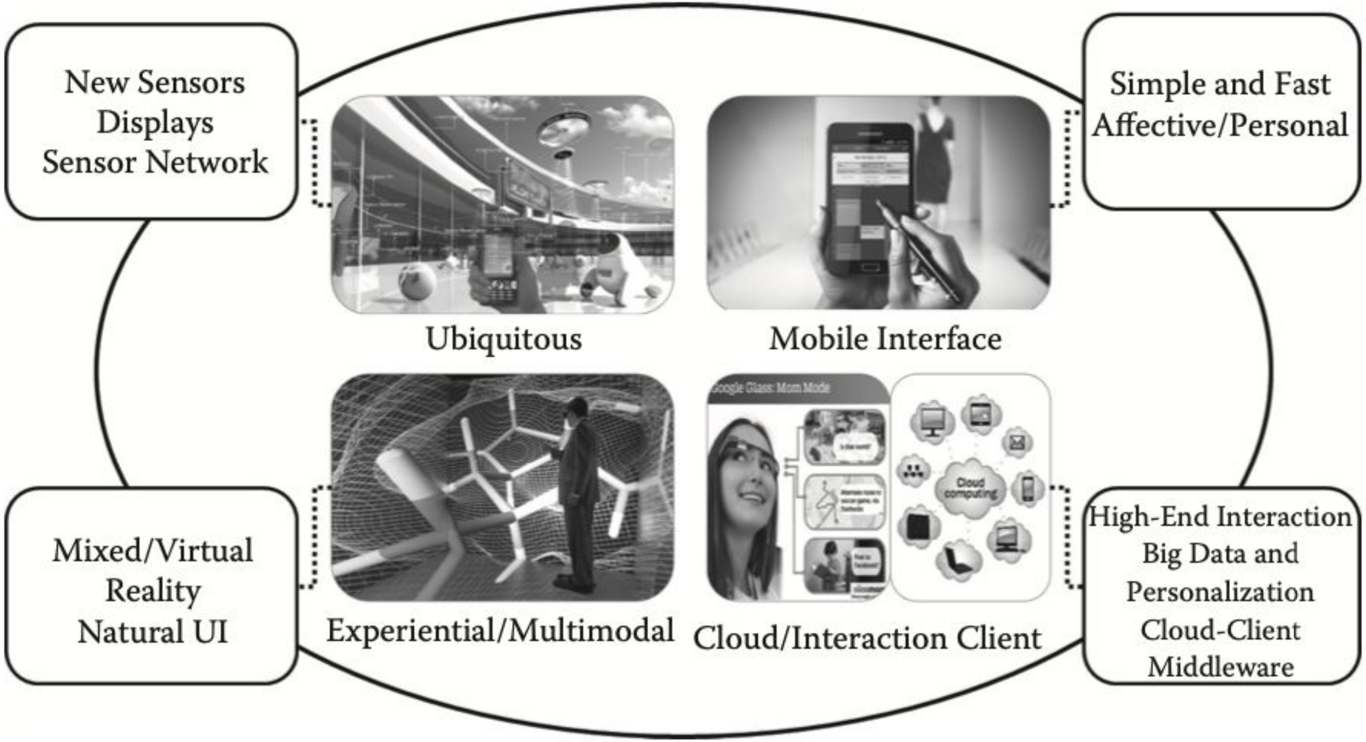
Gerard Jounghyun Kim

Chapter: 9
Future of HCI

Mahfuzur Rahman Emon
Lecturer, IICT, SUST

Four major computing platforms that have emerged in the past years:

- Mobile and handheld platform: (exemplified by the smartphones) which we can carry around to compute and communicate.
- Ubiquitous platform: In which everyday objects are embedded with interactive computing/networking devices and services
- Natural and immersive computing/sensing/display platform: That provides near-realistic services and experiences
- Cloud computing platform: That provides high-quality interactive services (based on its heavy-duty ultraserver level computing power) with real-time response (based on the fast network service).



Four emerging computing platforms and associated HCI technologies to pay attention to in the next 10 years: high-quality cloud service and ubiquitous and mobile interaction clients experiential and natural user interfaces.

Cloud computing platforms allow users to access applications through client devices like desktops and mobile devices. These devices lack the power for high-end services like image recognition and language understanding, so the cloud handles the heavy lifting. The cloud serves as the Model, while client devices act as the View/Controller. This setup enhances user experience by providing real-time, high-quality services tailored to different types of clients. Middleware solutions are needed to ensure seamless connection between the Model and client View/Controllers.

Mobile and ubiquitous platforms are driving the integration of embedded computers and sensors into everyday objects, known as the Internet of Things. Touch technology, evolving to include multi touch and proximity sensing, dominates interaction with these devices.

The interaction styles of the mobile/embedded vs. natural/realistic interfaces can be understood in terms of people’s natural dichotomous desires: one for simple and fast operations in a dynamic environment and the other for a rich and experiential interaction in a more stable, relaxed environment. These two desires are in tune with the lifestyle in the coming ages as we become more affluent and culturally richer. Virtual and mixed reality, multimodal interfaces are in the forefront of the experiential interaction technologies.

Interfaces are becoming more emotional, focusing on aesthetics, personalization, and adapting to users' changing needs. Achieving this requires advanced technology to sense contexts, user emotions, and intent, which is challenging even for humans. However, advancements in machine intelligence, like IBM Watson beating human champions, show promising progress.

In the following sections, we take a closer look at these promising HCI technologies, many of which are in an active stage of research

Non-WIMP/Natural/Multimodal Interfaces:

Considering various requirements and constraints, we found limited interface choices due to practical computing platform limitations (e.g., WIMP for desktop, touch-based for smartphones). However, as future computing platforms develop, we anticipate more options, including non-WIMP interfaces for natural and multimodal interactions.

Non-WIMP interfaces have been limited by issues like robustness and computational demands, but technological advancements and cloud computing are changing this landscape. We'll now assess future HCI technologies, including language understanding, gesture recognition, image recognition, and multimodal interaction.

Language Understanding:

The talking computer interface is undoubtedly the holy grail of HCI. Language understanding can be largely divided into two processes. The first is recognizing the individual words, and the second is making sense out of the sentence, which is composed of a sequence of recognized words

The current state of the art seems to be:

- Over 95% recognition rate (individual words)
- At least millions of words and more than 30 languages
- In real time through the high-performance cloud
- Without speaker-specific training
- In a mid-level noisy environment

Voice recognition technology is advanced enough for widespread use but is particularly valuable in situations of disability support or when both hands are occupied, not everywhere, Reasons could be

- Users are less tolerant to the 2%–3% of incorrect recognition performance
- Segmentation problem (quite difficult to separate and segregate the actual voice input from the rest ‘noise, normal conversation’ within the stream of voice
- Extracting the meaning from the words is another problem though the word level recognition is good enough.

Efforts are underway to improve voice/language understanding, exemplified by Apple Siri and IBM Watson. These technologies show promise for the future despite current limitations.

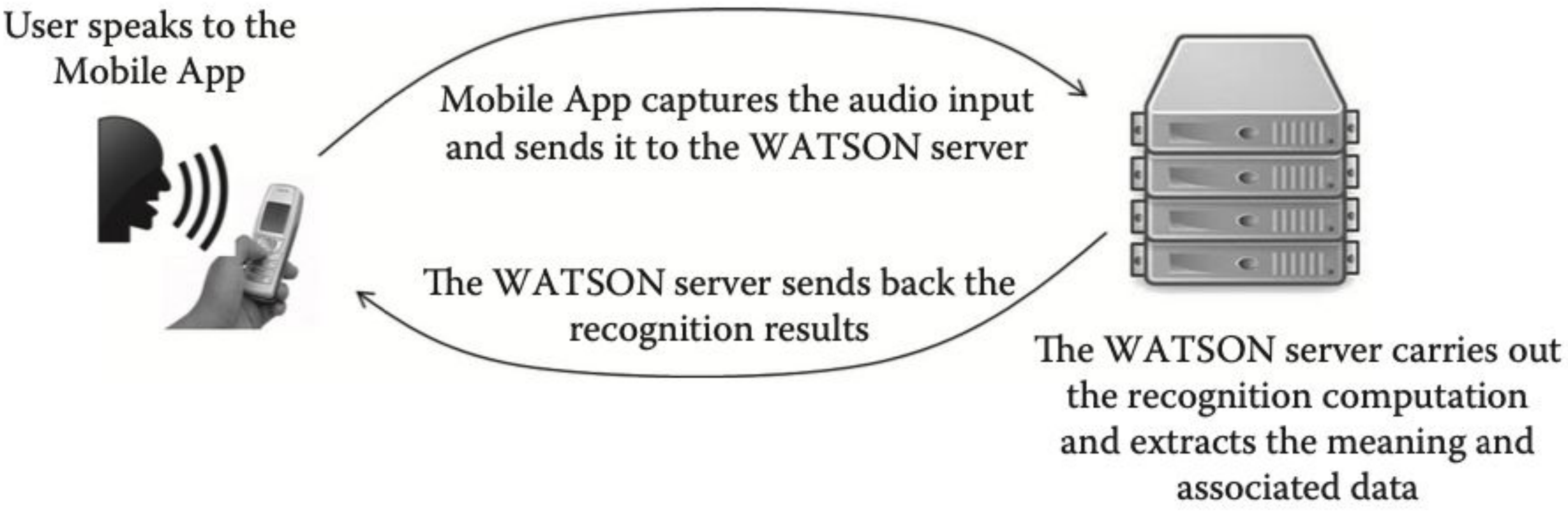


Figure 9.3 Voice/language-understanding service by the AT&T Watson cloud engine. (From AT&T Labs Research, http://www.research.att.com/articles/featured_stories/2012_07/201207_WATSON_API_announce.html?fbid=RexEym_weSd.)

Gestures:

Gestures alone can convey meaning, or they can function in a supplemental role in other modes of communication. the objective of incorporating gestures into human computer interaction is a natural outcome.

Gestures can be different types:

- From human's perspective (e.g. supplementary pointing vs symbolic)
  Supplementary pointing" refers to gestures that assist in indicating or emphasizing something, often in conjunction with verbal communication. For example, pointing to a specific object while discussing it.

  "Symbolic gestures" are non-verbal movements that represent abstract concepts or ideas. These gestures convey meaning without necessarily pointing to specific objects or locations. For instance, waving goodbye or making a thumbs-up gesture.

- Technological viewpoint (e.g. static posture vs moving hand gestures)
  A static posture refers to a gesture where the hand or body remains still without movement. It involves holding a particular position without changing or transitioning to another gesture. For example, holding one's hand up with fingers extended to signal "stop" or "wait" without any movement.

Hands/arms are used often for supplementary gestures (e.g., pointing) in verbal communication. For the hearing-impaired, the hands are used to express sign language.

Whether it is a static posture or involves movement of limb(s), must be captured over time. This is generally called motion tracking and can involve a variety of sensors that are targeted for many different body parts

Two-dimensional (2-D) hand/finger tracking are the ones using the mouse and touch screen. [direct contact with the device]

- In the case of the mouse, the user has to hold the device, and this is a source of nuisance, especially if the user is to express 2-D gestures rather than just using it freely to control the position of the cursor. This explains why mouse-driven 2-D gestures have not been accepted by users

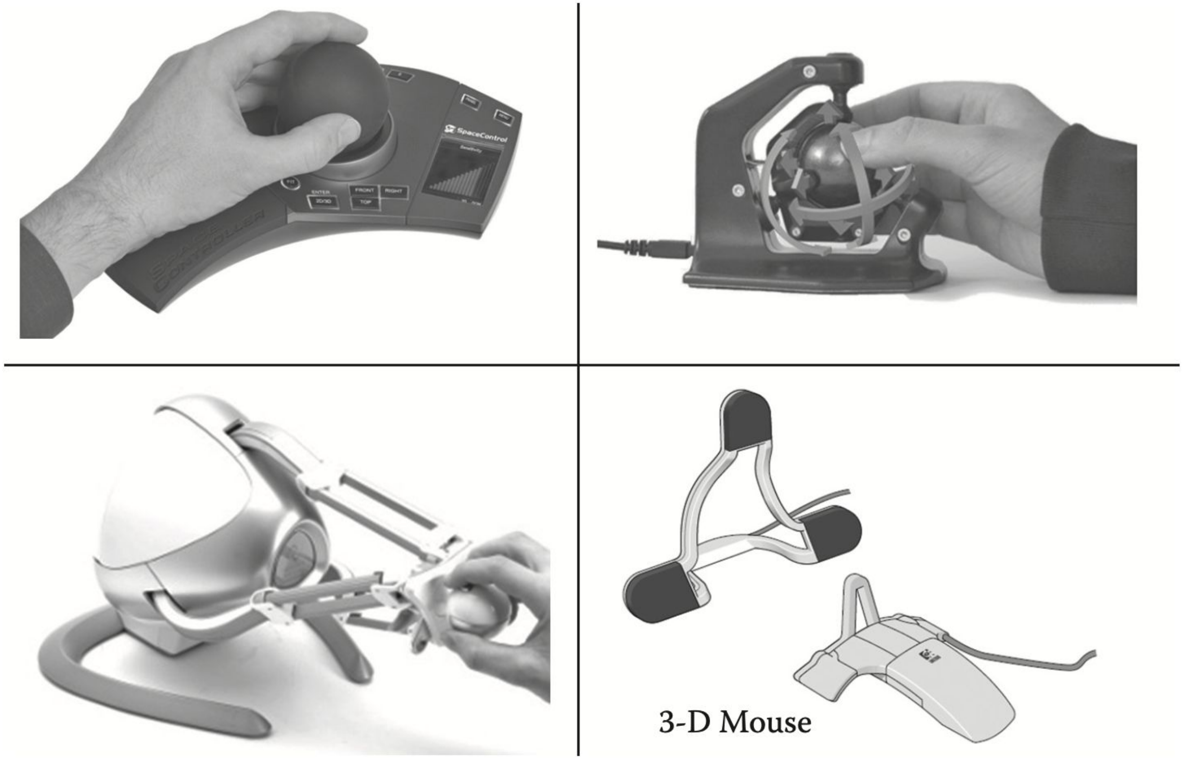
- Simple 2-D gestures on the touch screen, such as swipes and flicks, are quite popular.

- With the rise of small and embedded computing devices, traditional 2-D touch input may not always be feasible due to space constraints.

- Understanding gestures in three-dimensional space, similar to real-life movements, becomes crucial. This involves tracking the 3-D motion of body parts or objects.

- **Inside-Out Method:** Requires sensors to be held or attached to the body or object being tracked (e.g., 3-D mouse, Wii-mote). Various mechanisms such as electromagnetic wave detection or inertial sensors are used. Inconvenience in some case due to the need for physical attachment.

- **Outside-In Method:** Involves installing sensors external to the user's body (e.g., cameras or depth sensors like Microsoft Kinect). Users are free of attached devices, making movements feel more natural. However, tracking accuracy may be lower compared to inside-out methods due to the sensors being remote.



Examples of inside-out type (handheld) of sensors

- Camera-based tracking has become increasingly attractive due to advancements in computer vision technologies and algorithms. These advancements include improved accuracy, faster processing speed, and lowered costs.

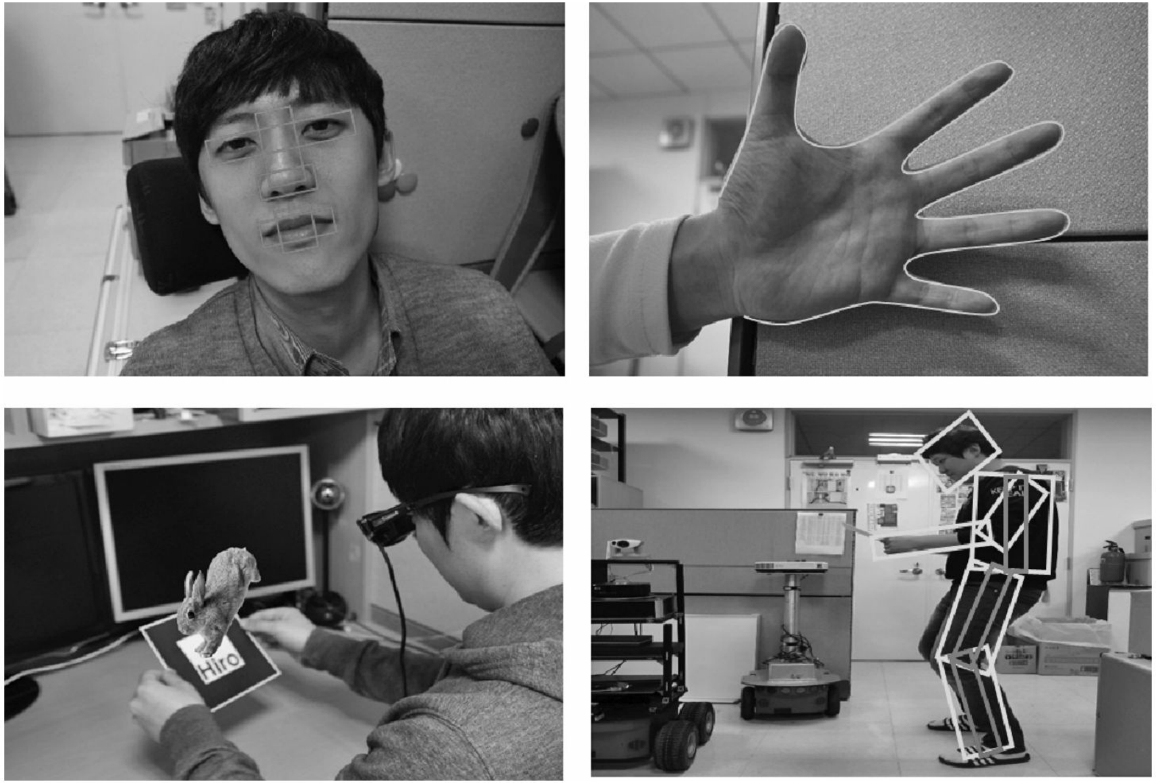
- Virtually all modern devices such as smartphones, laptops, desktops, and smart TVs are equipped with high-quality cameras, making camera-based tracking widely accessible.

- The increasing processing power of CPUs, GPUs, and multimedia processing chips enables more sophisticated tracking algorithms to be run efficiently.

- Standard and free computer vision, object recognition, and motion tracking libraries such as OpenCV and OpenNI make it easier for developers to implement camera-based tracking.

- Camera-based tracking performance can be affected by environmental factors such as lighting conditions.

- To ensure robust tracking, markers such as high-contrast geometric patterns, colored objects, or infrared LEDs may be used. This approach somewhat resembles the inside-out method as it involves attaching or using external objects for tracking.



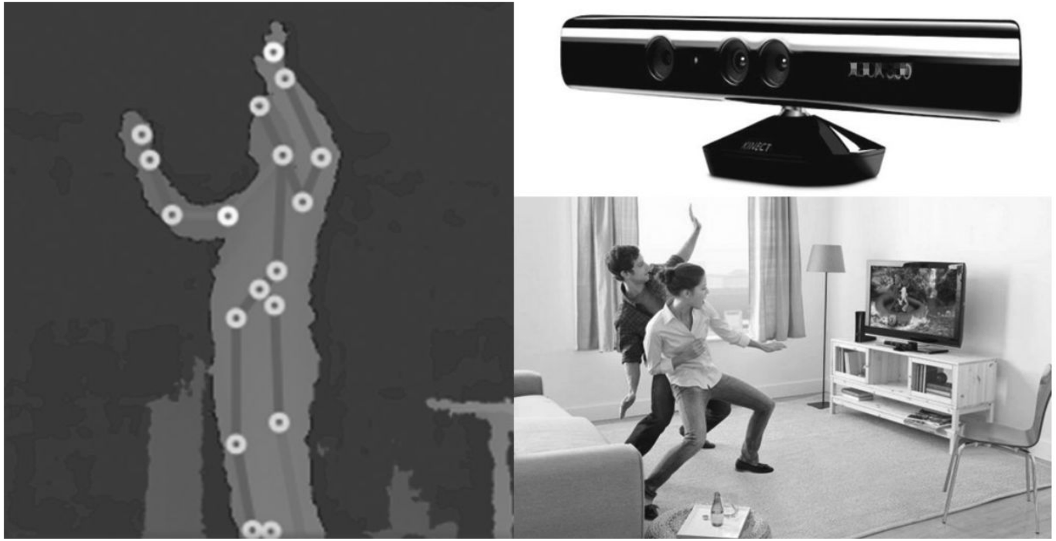
Camera-based motion-tracking examples

- Recent introduction of low-cost depth sensors has transformed the effectiveness, reliability, and practicality of outside-in gesture and motion-based interactions.

- The Kinect platform, utilizing both color cameras and depth sensors (originally developed by PrimeSense), can track whole-body skeleton motion of multiple users without requiring any devices to be worn. Originally designed for motion-based gaming, its applications have expanded to include environment reconstruction, motion capture, and more.

- Beyond gaming, Kinect-like technologies are utilized for environment scanning to generate computer models, motion capture, and various other applications.

- There's been development of smaller, miniaturized prototypes with comparable resolution and performance suitable for integration into mobile devices, enabling similar functionalities on a smaller scale.



Whole-body skeletonal tracking using the Kinect depth sensor (left) and its application to motion-based games (right).

Prototype miniature depth sensor mountable on mobile devices.

- Similar to the segmentation challenge in voice recognition, motion-based interaction faces difficulty in distinguishing meaningful gestures from continuous motion tracking data.

- Many existing systems rely on specific modes or states to operate, such as pressing a button to activate gesture recognition. However, this undermines the seamless, natural interaction intended with bare hand and outside-in sensing, reducing usability.

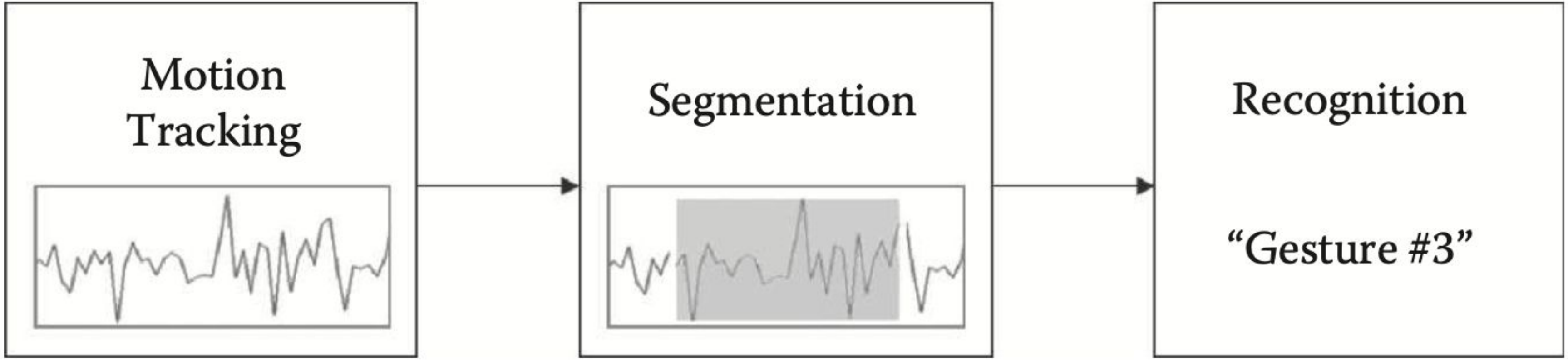
- New approaches, like those using "sliding windows" to continuously monitor motion streams for meaningful gestures, are being explored to address the segmentation problem.

- Unlike voice recognition, where background noise may be minimal and spoken inputs intermittent, motion gestures often involve continuous movement, making it challenging to extract gestural commands.

- Touch gesture is the same, it is natural to expect touches only when a command is actually needed. Thus a touch simply signals the start of the gesture input mode

- Combining multiple modes of interaction can help mitigate the segmentation problem by providing additional contextual cues.

- While motion-based interaction offers experiential and realistic experiences, it can also be physically tiring for users, highlighting the importance of balancing usability with functionality.

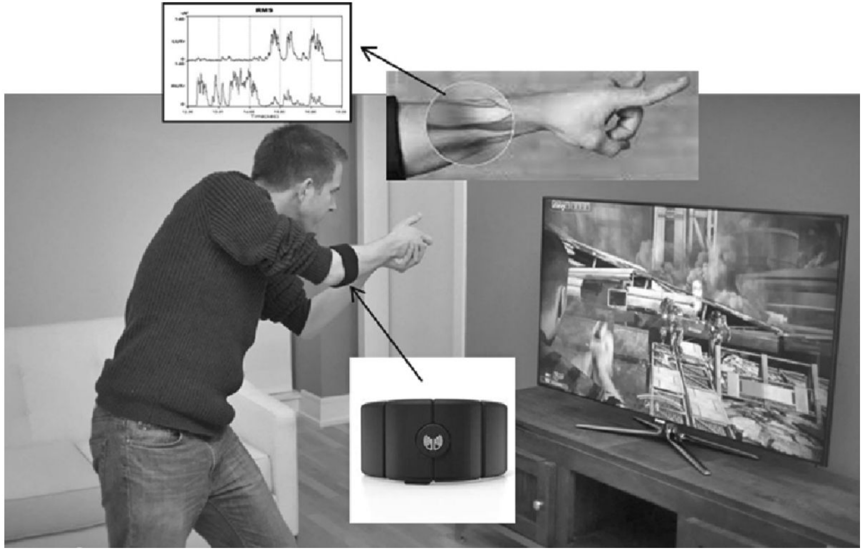


Three major steps in gesture recognition: (1) motion tracking, (2) segmentation (using the monitoring through the "sliding window" into the tracking data stream), and (3) recognition given the tracking data segment.

The sliding windows technique is a method used in pattern recognition to analyze sequential data by continuously monitoring a fixed or variable length of the data stream to detect meaningful patterns or gestures. It involves moving a window along the data stream and examining the data within the window to identify patterns or gestures.

- Detecting subtle finger articulations is difficult due to current sensor resolutions and the small size of fingers compared to the human body.
- Advances in sensor development and decreasing costs are mitigating this issue. Specialized depth sensors for finger tracking, like Leap Motion, are emerging in the market.
- Finger tracking was previously handled with glove-type sensors, but these were cumbersome and had low usability.
- Despite advancements, the usefulness of finger-based interaction in enhancing user experience (UX) is uncertain. In everyday life, fingers are primarily used for grasping rather than gestures, except in sign language. Even for touch-screen interaction, the number of finger-touch gestures is relatively small (e.g., swipe, flick, pinch).
- Electromyogram (EMG) sensors are being explored for recognizing motion gestures by approximating joint movement. This technology could offer new possibilities for gesture recognition in gaming and other applications.

Finger-based interaction using the Leap Motion.



Wristband type of EMG sensor for simple gesture recognition

Electromyography (EMG) measures muscle response or electrical activity in response to a nerve's stimulation of the muscle.