11 User Interface and Future Interaction Technologies

11.1 Introduction: The shift towards next generation user interfaces e.g., psychology, sociology and anthropology

Consumers around the world are now able to establish and maintain connection, regardless of location, through smart mobile devices. Coupled with a wide variety of applications and cloud service offerings, this has generated new avenues through which consumers access content and has also triggered shifts in consumer behaviour. For instance, with consumers now using multiple smart electronic devices for entertainment (e.g., video streaming), the proportion of consumers watching TV broadcasts and cable has dropped drastically. Revenue from digital streaming in the USA was up 545% from US$85 million in Q1 2011 to US$529 million in Q1 2012[1], replacing part of cable TV sales.

Similarly, the increased use of apps and cloud services has led to a reduction in purchases and rentals of DVDs. Total US DVD rentals in Q1 2012 fell approximately 18% to around US$1.8 billion while rentals from stores dropped more than 39%. Concurrently, user interface and interaction technologies will evolve to match consumer expectations of a more seamless, intuitive and immersive user experience.

These expectations bring new opportunities foruser interaction with data, systems and the environments in which they work, live and play. Figure 1 shows the key aspects of how user interfaces and interaction technologies have evolved over the past 50 years: specifically, an increase in size and number of units sold, improvement in degree of interactivity (from static batch processing to intuitive interfaces) and degree of integration with other devices and technologies. This evolution marks the beginning of a transformational shift, with the user experience through computing form factors superseding the effects of product functionalities.

[1] Michael Cieply. *Report Shows Quarterly Decline in Video Rental Revenue; Digital Streaming Increases.* [Online] Available from: http://mediadecoder.blogs.nytimes.com/2012/04/29/report-shows-quarterly-decline-in-video-rental-revenue-digital-streaming-increases/?nl=business&emc=edit_at_20120430 [Accessed 9th July 2012].

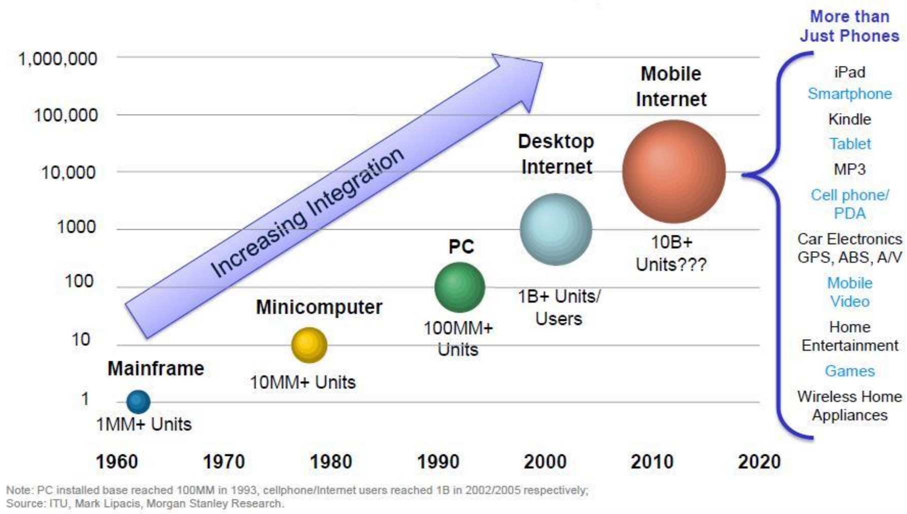

Figure 1: Evolution of the User Interface

We have witnessed certain events that are testament to the transformational shift. Traditional mobile device incumbents such as Nokia and Ericsson have lost their dominance in the consumer market and a new generation of intuitive computing devices, pioneered by key players such as Apple and Samsung, has emerged. It is common knowledge that there exists a strong market demand for Apple's smart mobile devices and the convergence of multiple intuitive features such as multi-touch, speech recognition, natural language question and answer capabilities on its devices. For example, Apple 4S Siri app, with its ability to answer simple queries such as the weather and word definitions, is recognised as a first step toward artificial intelligence capabilities in mobile devices.

Ultrabooks, which are new thin and light PC designs will emerge, incorporating tablet design features into PCs. Tablets could also become future input/output interfaces for other electronic devices.

The epitome of the user interface now lies in improving the user experience, defined as a user's perceptions and responses that result from the use or anticipated use of a product, system or service[2]. The economic opportunities presented by the user interface are by no means trivial. The size of the global user interface market is predicted to grow from US$10 billion in 2011 to US$25 billion by 2016[3].

User interface and interaction technologies are relevant to the future of Singapore. IDC projected figures of smartphone and tablet shipments suggest the possibility of a smartphone penetration of nearly 100% and a tablet adoption of approximately 30% by 2013. In addition to a sophisticated ICT-savvy population, these trends will lay the stage for

---

2 Iso.org. Ergonomics of human system interaction – Part 210: Human-centred design for interactive systems. [Online] Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075 [Accessed 9th July 2012].

3 PRNewswire. *Mobile User Interface 2011-2016: Designing Compelling UI in iPhone and Android Era.* [Online] Available from: http://www.prnewswire.com/news-releases/mobile-user-interface-2011-2016-designing-compelling-ui-in-iphone-and-android-era-130548743.html [Accessed 9th July 2012].

further adoption of user interface and interaction technologies as well as relevant market opportunities in Singapore.

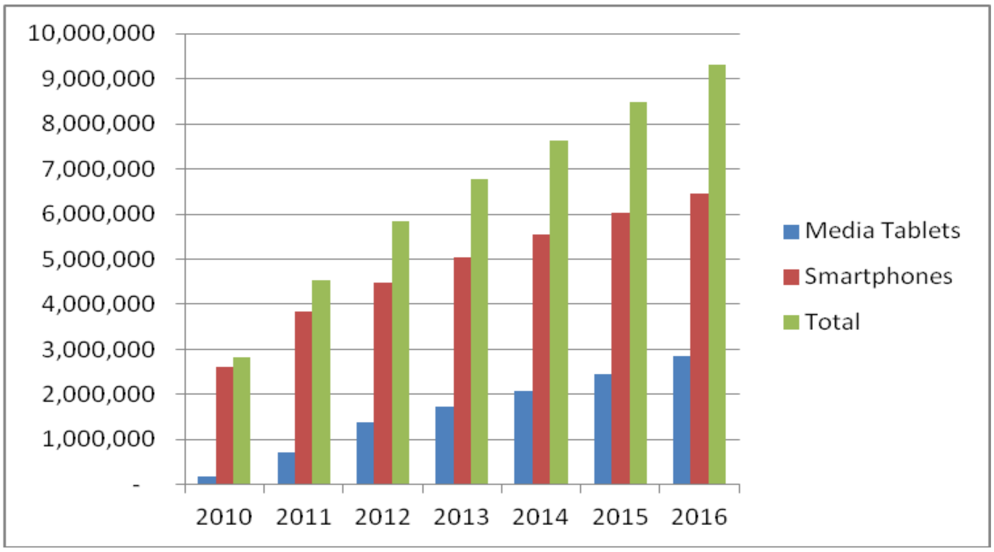

Figure 2: Projected Shipments of Smartphone and Tablet to Singapore from 2010 to 2016. (Source: IDC[4])

11.2 Market Trends

Three market trends provide the main drivers of change for user interface and interaction technologies. They are namely visual info-gratification, the user as the new interface, and smarter devices and emerging user interfaces for an immersive user experience.

Visual info-gratification refers to consumers accessing information about real-world objects, naturally and on-the-go, simply by pointing their smart devices at an object of interest. There are already existing consumer apps that provide value-added services through visual info-gratification. French site WhereToGet.It allows users to post photos from the street, magazines, blogs or films, and ask where the featured items can be purchased. Amsterdam's Schiphol Airport and Paris' Charles de Gaulle Airport launched a mobile app in January 2012 specifically designed to help Chinese visitors navigate around the airports, by translating the airport signs in real-time. VizWiz is an app that blind users "see", by providing them quick answers to questions about their surroundings through images they have taken. Google's Skymap enables users to point their phones at the sky to display detailed information on the objects or constellations that are pointed at.

The concept of natural user interface has also shattered the existing paradigm of user interfaces by positioning users as the new user interface. Microsoft Kinect was developed, based on this concept, and is able to decipher images, gestures and speech commands from users. It sold about 18 million units in 2011 and attained the Guinness World Record for

[4] IDC. *IDC Asia/Pacific Quarterly Mobile Phone Tracker*, released in May 2012 and *IDC WW Quarterly Tablet Tracker*, released in May 2012

being the fastest selling consumer electronics device in history[5], even exceeding that of the iPhone and iPad. In early 2012, Microsoft released the commercial version of Kinect with far-range capabilities at an attractive price point of US$250 and made available the Software Development Kit for free, with the belief that the community would accelerate the pace of innovation based on Kinect's capabilities.

The Holy Grail for the consumer is an immersive and holistic user experience through the convergence of multiple user interface technologies. Microsoft's Ultimate Battlefield 3 Simulator, an immersive environment to simulate battle scenes, provides such a user experience. It is equipped with an omni-directional treadmill to enable movement across any distance in any direction in the highly photo-realistic virtual world. It is able to detect motion through infra-red and is also able to change the player's game perspective (be it standing or squatting) on demand. The player is able to manipulate weapons through speech commands and gestures and "feel" the pain if shot during battle.

Although the convergence of multiple user interface technologies for immersive environments was developed in specific industrial contexts (e.g., maintenance training and familiarisation) and is not new *per se*, the application of user interface technologies for immersive environments will eventually extend from the industrial to the consumer context.

To keep pace with rising consumer expectations, technologies are evolving to enhance the user experience. For example, surface computers (computers that interact with users through the surface of ordinary objects rather than through a monitor and keyboard) have emerged with more advanced multi-touch capabilities, and more apps and services developed for multi-user collaboration. Microsoft launched the second generation surface computer, the Samsung SUR40, this year while Samsung released a new series of smart televisions with voice and gesture recognition.

In late 2011, Apple filed a patent to incorporate Kinect-like motion in future iPhones, revealing an exciting 3D user interface for future devices. The new interface could possibly include proximity sensor arrays, enabling the smart device to respond to hovering gestures. Beyond multi-touch and gesture recognition, other user interface capabilities such as natural language question and answering systems (providing users with a means of asking a question in plain language and receiving a meaningful response), facial recognition and eye tracking (monitoring facial expressions and eye gazes) and speech-to-speech translation (enabling the user to easily understand any given speech phrase in real time) may find their way into future smart devices.

The user interface is also an active area for ICT R&D. Intel has established a dedicated Interaction and Experience R&D group to shape new user experiences and its future computing platforms[6]. The European Union is also embarking on an R&D track to improve

---

5 Guinness World Records. *Kinect Confirmed As Fastest-Selling Consumer Electronics Device.* [Online] Available from: http://community.guinnessworldrecords.com/_Kinect-Confirmed-As-Fastest-Selling-Consumer-Electronics-Device/blog/3376939/7691.html [Accessed 9th July 2012].

6 Electronista. *Intel starts up dedicated user experience R&D group.* [Online] Available from: http://www.electronista.com/articles/10/06/30/intel.launched.ixr.division.for.user.interfaces/ [Accessed 9th July 2012].

immersive and interactive media technologies, providing users with more sophisticated forms of media and enhanced experiences[7].

11.3 Technology Outlook

Enabling technologies in the different time frames are described as follow. (Figure 3 provides an overview of the enabling technologies on a timeline.)

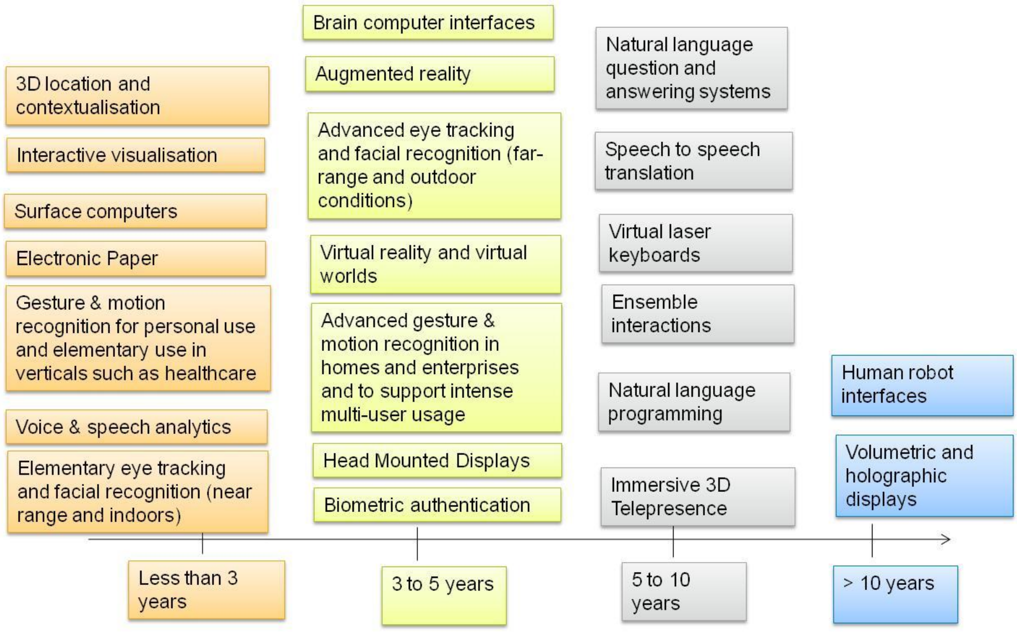

Figure 3: Timeline of enabling technologies

11.3.1 Less than three years

Electronic (e-)paper

Electronic or e-paper refers to reflective display technologies that do not require a backlight and can be viewed in conditions of moderate to good ambient illumination. DisplaySearch forecasts that the total e-paper display units sold will grow from 22 million units in 2009 to 1.8 billion units in 2018 at a compound annual growth rate (CAGR) of 64% while the total e-paper display market size will expand from US$431 million in 2009 to US$9.6 billion in 2018, growing at a CAGR of 41%[8].

7 European Commission. *Updated Work Programme 2011 and Work Programme 2012: ICT - Information and Communications Technologies.*; 2011.

8 DisplaySearch. *E-Paper Display Revenues Forecast to Reach $9.6 Billion by 2018, DisplaySearch Reports; E-Books to Reach 77M Units by 2018.* [Online] Available from:

E-paper can be made very thin, producing a nearly paper-thin rewritable display that gives a similar user experience to that of printed paper. Unlike conventional cathode ray tube (CRT) or backlit flat panel displays (thin film transistor - TFT, liquid crystal display - LCD), e-paper is often considered to be more comfortable to look at, compared to the majority of common display monitors. This is because of its stable image which does not need to be refreshed constantly, a wider viewing angle and the fact that it reflects ambient light instead of emitting its own. Touch sensitivity can be added to electronic paper by incorporating a touch layer, enabling features such as the highlighting of words or adding of handwritten notes in electronic or e-books.

A promising development of e-paper is the concept of electrowetting (EWD), which relies on controlling the shape of a confined water/oil interface with an applied voltage. Electrowetting will enable e-paper to display video content through fast switching between white and coloured reflection. This adds a new dimension to e-books, exposing the reader to richer video information on demand while maintaining the qualities of a book. As a result, content delivery ecosystems can be transformed with potentially more collaborations between authors, digital content producers and device manufacturers, leading to new content delivery business models.

In May 2011, researchers at Queen's University in Ontario Canada unveiled a smartphone that had a flexible e-paper display instead of an LCD/TFT touchscreen[9]. The device is a fully functional phone, complete with voice and texting capabilities and is thin enough to fit in a wallet. Lying beneath the thin-film electronic ink display is a printed circuit board that contains resistive bend sensors, which can be programmed to determine various bending gestures. This allows users to bend the top corner to turn a page or navigate a menu. In addition, the smartphone has a good battery life as it does not consume electricity when switched off and the e-paper display consumes very little, even with intense usage. Similarly, Hewlett Packard (HP) and Arizona State University collaborated to develop an affordable e-paper display. Although the prototype is monochrome, they hope to launch field trials of a colour-flexible display within three years.

---

http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/090826_e_paper_display_revenues_forecast_reach_9_6b_by_2018.asp [Accessed 9th July 2012].

9 diTii. *'PaperPhone': Flat, Flexible e-Paper Display - Smartphone and Tablet Unveiled.* [Online] Available from: http://www.ditii.com/2011/05/06/paperphone-flat-flexible-e-paper-display-smartphone-and-tablet-unveiled/ [Accessed 9th July 2012].

Figure 4: Prototype of e-paper smartphone[10]

The initial major applications for e-paper are el-books, signage (in retail and roadside applications), and small information-centric screens (in mobile phones and music players). A future application may be in automotive dashboards, using the high contrast ratio of e-paper.

Interactive visualisation

Interactive visualisation technology displays data using interactive images with the colour, brightness, size, shape and motion of visual objects representing aspects of the dataset being analysed. These products provide an array of visualisation options including heat maps, geographic maps, scatter plots and other special-purpose visuals. These tools enable users to extract insights from the data by interacting with its visual representation. Interactive visualisation is poised to grow as enterprises adopt an "information-driven approach".

Key barriers to mainstream adoption are mainly the lack of integration with the business intelligence capabilities of enterprises and the lack of knowledge of how to exploit this technology for business value. However, this can change with vendors offering solutions that are integrated with their business analytics capabilities. In March 2011, MicroStrategy launched its data visualisation tool, Visual Insight, integrated with its MicroStrategy 9.2 business analytics solution. This allows users to access business data modelling and interact freely with the data without being reliant on IT departments and vendors[11]. This represents a shift toward more user-centric data analysis capabilities from traditional IT-centric and report-centric approaches. There are also more solutions that incorporate interactive visualisation, namely Adobe Flash, Microsoft Silverlight, Ajax, HTML5 and other Web 2.0 technologies that enable animated, interactive displays of data.

Interactive visualisation can be used in various sectors. In security, visualisation of cybersecurity data may uncover hidden patterns, and identify emerging vulnerabilities and attacks, enabling a quick response with countermeasures. In finance, interactive visualisation can facilitate a quick overview of enterprises' financial performance, helping potential investors and lenders identify credit risk and opportunities. In healthcare, an intuitive visualisation of patients' health indicators,instead of the usual review of multiple disparate medical reports, helps healthcare practitioners identify anomalies quickly. These examples illustrate the insights that users can obtain by interacting with the visual representation of data.

Today, several governments (e.g., the USA, UK and Singapore) are opening up their datasets for the developer community to develop interactive visualisation applications to extract

---

10 HumanMediaLab. *Paper computer shows flexible future for smartphones and tablets.* [www.youtube.com] .

11 MicroStrategy. *MicroStrategy Gives Business Users Visual Insight.* [Online] Available from: https://resource.microstrategy.com/ResourceCenter/transmit.aspx [Accessed 9th July 2012].

further insights from available government data. Such applications are equipped with interactive visualisation features for analysts to obtain trends in a more intuitive manner. Some examples from the US Government Open Data Initiative include the Centres for Medicare and Medicaid Services' (CMS) Inpatient Dashboard (to track Medicare in-hospital spending) as well as the USA Spending Dashboard on Government Spending (to provide trends on Government spending).

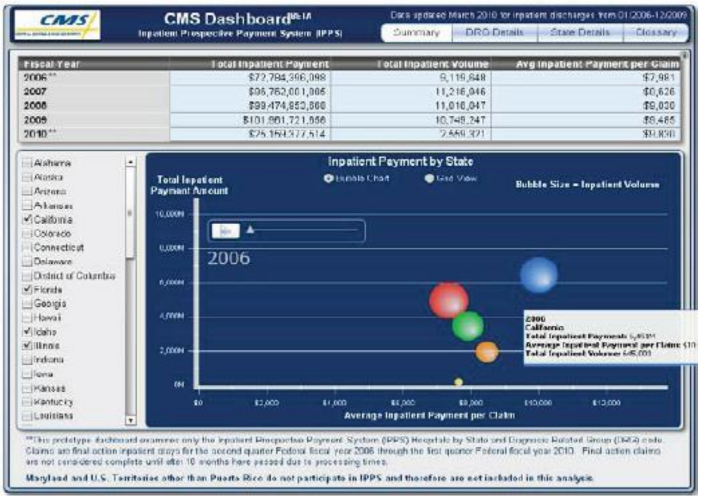

Figure 5: The application for Centre for Medicare and Medicaid Services' (CMS) Early Warning System to track effectiveness of Medicare fraud prevention efforts[12]

Seamless Indoor and Outdoor Location Positioning and Contextualisation

A person's location is instrumental in determining the context of the individual in the overall user experience. Location determination technologies have been improving and today's capabilities can give an accuracy of up to a few metres and even in indoor environments. The common location determination technologies that provide location positioning are cellular-based (e.g., GSM, 3G, 4G), Wi-Fi and Global Positioning Systems (GPS).

Two-dimensional (2D) or 3-dimensional (3D) map data provide raw location coordinates of the user and give a visualisation of the user's location and the available facilities in the vicinity. Today, 2D geospatial data and location-based services such as car navigation and location sharing on social media websites are already in place. However, the challenge is to enable seamless indoor and outdoor location positioning and contextualisation for more sophisticated location-based services. Some ongoing challenges are the accuracy of indoor location determination technologies indoors, availability of indoor building maps (which could be addressed by recreating indoor environments from scratch) and the computational power required to display outdoor and indoor 3D environments on mobile devices.

12 HHS.gov. *New High-Value Data Sets and Tools.* [Online] Available from: http://www.hhs.gov/open/plan/opengovernmentplan/transparency/dataset.html [Accessed 9th July 2012].

Other contextual information is not limited to the physical world around the user, but also incorporates behavioural and social network characteristics. Individuals can have a "digital identity", taking on various personas from work to leisure as well as accompanying devices and communication methods to stay connected. This "digital identity" can be broadly defined under a contextual information taxonomy of 'What', 'Who', 'When', 'Why' and 'Where'. For example, by snapping a photo of a historical monument and knowing that the user is a secondary school history student, the mobile device can retrieve related information from the secondary school history syllabus, tourist-related information and useful services in the vicinity for the student. In addition, fusing the location data of users, who the users are with and the time of query could unveil the context that the users are in, rendering appropriate information and services more accessible through the mobile device.

Taxonomy	Examples of contextual information available
'What'	▪ User geospatial motion information (e.g., orientation, speed, acceleration) ▪ Environmental information (e.g., temperature, air quality) ▪ Physiological measurements (e.g., blood pressure, heart rate, respiration)
'Who'	▪ Social contacts nearby ▪ Social networks of similar interest to the user
'When'	▪ Calendar of events ▪ Temporal information (e.g., time of the day, date, season of the year)
'Why'	▪ Inferred activities (e.g., running, shopping for a product) ▪ Profiling of consumers based on their activities
'Where'	▪ Location coordinates in x, y and z dimensions ▪ Position of coordinates on 2D and 3D maps

Table 1: Contextual Information Taxonomy[13]

Surface Computers

13 Accenture. Accenture Technology Vision 2012. [Online] Available from: http://www.accenture.com/us-en/blogs/accenture-blog-on-insurance/Media/ACN_Technology_Vision_2012.pdf [Accessed 9th July 2012].

Surface computers are large-screen displays that support direct interaction via multi-touch or gesture. The displays can typically recognise multi-touch, enabling multiple users to interact or work collaboratively. Microsoft has released the second generation of its surface computers (Microsoft Surface using Samsung SUR40 display, now called the Samsung SUR40 with Microsoft PixelSense) this year at a price point of US$8,400. Surface computers are currently targeted at the hospitality and retail industries. With price points expected to go down over the next few years, and more apps and services developed on the surface computer platform, surface computers are expected to see mainstream adoption in less than three years.

Gesture Recognition and Motion Sensing for Elementary Use

Gesture recognition involves determining the movement of a user's fingers, hands, arms, head or body in three dimensions through the use of a camera or a device with embedded sensors that may be worn, held or body-mounted. With Microsoft making the Software Development Kit for the commercial version of Kinect free, developers will be encouraged to innovate and create apps in multiple areas, including business, healthcare and gaming. In the short term, elementary gesture recognition technology such as the manipulation and retrieval of data on smart and mobile devices, will see mainstream adoption.

In addition, motion detection capabilities are already on smart devices today. Micro-electro-mechanical systems (MEMS) enable a multitude of applications from motion detection to navigation. Mobile phones today already have MEMS sensors such as accelerometers (rotation of phone display according to user movement), gyroscopes (maintain or measure orientation), barometers and altimeters (measure atmospheric pressure) and magnetometers (provide compass direction functionality). The global motion sensor market is slated to reach US$16.4 billion in 2016, with a CAGR growth of 20.3% from 2011 to 2016 while the motion sensor in mobile phones will reach US$4.1 billion, with a CAGR growth of 25.3% from 2011 to 2016[14] (see Figure 6).

---

[14] PRNewswire. *Mobile User Interface 2011-2016: Designing Compelling UI in iPhone and Android Era.* [Online] Available from: http://www.prnewswire.com/news-releases/mobile-user-interface-2011-2016-designing-compelling-ui-in-iphone-and-android-era-130548743.html [Accessed 9th July 2012].

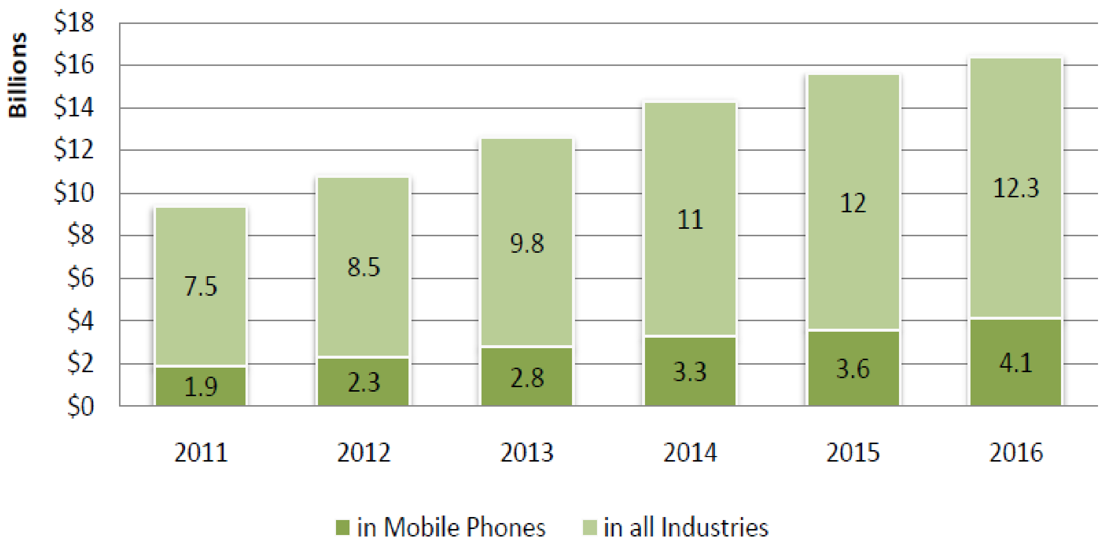

Figure 6: Global Motion Sensing Market from 2011 to 2016[15]

Elementary Eye Tracking and Facial Recognition

Eye tracking technologies determine the angle or position of a user's visual attention, typically through the use of cameras. The different types of non-intrusive eye trackers generally include two common components: a light source and a camera. The light source (usually infrared) is directed toward the eye while the camera tracks the reflection of the light source. The data obtained is used to extrapolate the rotation of the eye and derive the direction of gaze.

Eye tracking can provide the following forms of data for analysis. Scan paths allow researchers to analyse the flow of attention throughout a design. The scan paths show eye fixations and the order in which they occur. This provides insights to feature visibility and the visual impact on customers. Sequential videos can also show scan path progression in real-time. Heat maps display where users concentrate attention within a given design, revealing which areas people focus on, as well as areas that are commonly ignored. Figures 7a and 7b show examples of heat maps as well as scan paths.

15 PRNewswire. *Mobile User Interface 2011-2016: Designing Compelling UI in iPhone and Android Era.* [Online] Available from: http://www.prnewswire.com/news-releases/mobile-user-interface-2011-2016-designing-compelling-ui-in-iphone-and-android-era-130548743.html [Accessed 9th July 2012].

Figure 7a (left): Eye track heat map; **Figure 7b (right)**: Eye scan path[16]

Facial recognition, on the other hand, analyses facial features and expressions to infer more information – age, gender and emotional states - on the user. Cameras have been placed in billboards and advertisements to profile the type of viewer looking at the advertisement and enable the advertisement to change dynamically to suit the profile of the viewer. Face.com, an Israeli startup, has been offering a face-recognition service that websites or apps can use to count the number of faces in a photo, identify gender, or match the faces to known individuals. It can also estimate the age based on the face found in photographs and ad networks.

However, the use of facial recognition is still in its infancy. It only works best with images and videos taken in indoor and well-lit environments. A key challenge to improve accuracy is to teach the system to recognise certain lighting conditions that affect age estimation. For example, strong light from above accentuates shadows around a person's eyes and may lead to the system estimating an age much higher than the person's actual age.

Elementary Voice and Speech Analytics

Voice stress analytics is an elementary form of voice and speech analytics, which recognises a person's stress level by analysing the characteristics of the voice signal, with word use, if available, as an additional input. It is primarily used in call centres to identify angry or abusive customers and alert a supervisor when stress or anger is detected. As changes in stress levels are analysed, the content of the conversation becomes stationary.

When the content of the speech is understood, speech analytics can be used to answer simple questions and perform mundane operations. An example of this is Apple's Siri, which can process a user's speech in natural language, reply the user within a reasonable period of time and perform routine tasks. Using voice and speech analytics, iPhone 4S users can make reservations at specific restaurants, buy movie tickets or call a taxi by dictating instructions

16 Bentley University. *Design and Usability Center.* [Online] Available from: http://usability.bentley.edu/eye-tracking [Accessed 9th July 2012].

in natural language to Siri. Users can also pose simple queries such as "What is the weather tomorrow?"

Although the accuracy of speech analysis, relating to colloquial language use and accent issues is still an issue, these capabilities are rapidly evolving and in less than three years, the accuracy will be sufficient for mainstream adoption in smart devices.

11.3.2 Three to five years

Augmented reality

Augmented reality (AR) superimposes graphics, audio and other virtual enhancements over a live view of the real world. AR aims to enhance user interaction with the environment in real time through the digital environment. The concept of augmented reality is not new and has been used in various industrial contexts. In engine maintenance, engine diagnostics with data from sensors are superimposed on the real life view of the engine, enabling mechanics to identify which components are working and which are faulty. In vehicle assembly, the location of wire bundles and mounting points are superimposed on the real-life view of the truck chassis, allowing operators to more accurately place the wire bundles and affix them accordingly.

The maturity of a number of mobile technologies — such as GPS, cameras, accelerometers, digital compasses, broadband, image processing and face/object recognition software — has made AR a viable technology on mobile devices. In addition, there is an increasing software support for AR adoption on mobile devices. LG Electronics, in collaboration with augmented reality pioneer Wikitude, announced the adoption of the world's first 3D AR browser on its Optimus 3D smartphone[17]. More AR apps are being developed, with many of them available for free. Pinkfroot has developed Plane Finder iPhone and Android augmented reality apps, which allow users to point their phone cameras at the sky and discover individual aircraft flight numbers, speed, altitude and distance.

In December 2010, Word Lens launched an AR translation app allowing users to translate one language to another by pointing a camera at a traffic sign. Saltlux, a Korean company, has developed an AR mobile app that identifies restaurants by pointing the mobile phone camera at the place of interest, providing reputation analysis based on social media, social network recommendations and dynamic mobile social networks.

[17] Robert Nazarian. *Wikitude 3D Augmented Reality coming to LG Optimus 3D.* [Online] Available from: http://www.talkandroid.com/40352-wikitude-3d-augmented-reality-coming-to-lg-optimus-3d/ [Accessed 9th July 2012].

Figure 8: Plane Finder Augmented Reality app[18]

Despite the plethora of AR apps, some obstacles need to be addressed before mainstream adoption. Due to demanding AR requirements for mobile devices, use of AR will largely be limited to high-end smartphones and tablets. Mobile devices also have smaller screens than other consumer electronics devices such as laptops, restricting the amount and the layout of information that can be conveyed to the end user. There are also issues with compatibility. With competing AR browsers using proprietary APIs, AR information from one vendor's browser is incompatible with that from other browsers.

Advanced Gesture Recognition and Motion Detection

Current gesture recognition capabilities include analysing two or three users' gestures simultaneously, with gestural commands mainly used in the gaming context. Advanced gesture recognition will involve larger groups of users interacting at the same time as the system responds to these gestures within a shorter period of time. This is a challenge as considerable data processing is needed to recreate wire frame models of body positions and vector-based dynamics (for speed and direction of movement), followed by the interpretation of these gestures into meaningful commands for an application.

The creation of intuitive and logical natural user interfaces for home and business applications is another challenge. The conceptual design of a user interface based on gestures is a considerable task — not only from a technical standpoint but also from a cultural and anthropological perspective, especially in a global market where cultural sensitivity must be taken into account. Research-wise, the SixthSense project at the Massachusetts Institute of Technology (MIT) links the use of gesture recognition with AR to explore a new generation of interactions, for example, projecting more information about the physical objects that the user is interacting with.

In another research project, researchers at Microsoft and the University of Washington are working on harnessing electromagnetic radiation for a computer interface that turns any

18 pinkfroot. *Plane Finder AR Flight Tracker App.* [Online] Available from: http://my.pinkfroot.com/page/plane-finder-ar-track-live [Accessed 9th July 2012].

wall in a building into a touch-sensitive surface[19]. Harnessing electromagnetic radiation through gestures can allow light switches, thermostats, stereos, televisions, and security systems to be controlled from anywhere in the house, and can lead to new interfaces for smart homes.

Advanced Eye Tracking and Facial Recognition

Advanced eye tracking and facial recognition reasonably determine the facial expressions and eye gazes of multiple users from great distances at the same time and in dynamically changing environments such as different lighting conditions and with moving crowds. With such advanced capabilities, more interactive and effective advertising campaigns can be launched through fixed or mobile platforms and businesses will be more effective in delivering advertisements based on reasonably accurate consumer profiles. In homeland security, immigration and border authorities can use eye tracking and facial recognition to identify potential immigration offenders. In the UK, a sophisticated video camera system to detect emotions and lies has been developed by researchers at the universities of Bradford and Aberystwyth, in conjunction with the UK Border Agency[20]. The camera is able to detect visible signs in eye movements, dilated pupils, biting or pursing lips, wrinkling noses, heavy breathing, swallowing, blinking and facial asymmetry.

Virtual reality and virtual worlds

Virtual reality and virtual worlds provide computer-generated 3D environments that surround a user and respond naturally to his actions. This has already been applied to traditional niche areas of high-end simulation and training applications, including flight simulators, truck operator training in specialised environments (such as mines) and accident investigation in several industries.

Second Life, the user-generated virtual world, generates about US$75 million in revenue a year. Second Life is one of the more popular virtual worlds and has been around since 2003. Its usage statistics seem to be stable, with the company neither gaining nor losing significant numbers of users and reporting consistently strong economic growth. Virtual worlds have initially been thwarted by a lack of a clear audience value proposition across age groups and the growth of virtual worlds have been largely constrained to niche areas such as those focused on children and pre-teens (e.g., Habbo, Moshi Monsters, Stardoll and Club Penguin), education, virtual events, and simulation/training. With the dwindling interest in virtual worlds generally, virtual reality and virtual worlds could see mainstream adoption in about three to five years.

An emerging specific use of a virtual world is a mirror world for modelling and simulation, which may be a precise replication of the 3-D environment of a city, with a resolution 10

---

19 Kate Greene. *Talking to the Wall.* [Online] Available from: http://www.technologyreview.com/news/423905/talking-to-the-wall/ [Accessed 9th July 2012].

20 Hamish Pritchard. *New emotion detector can see when we're lying.* [Online] Available from: http://www.bbc.co.uk/news/science-environment-14900800 [Accessed 9th July 2012].

times higher than that of Google Earth but requiring roughly 100 times more data. Companies such as Microsoft, GeoSim Systems, Iwane and GElement have developed end-to-end solutions to recreate cities and indoor environments, drawing data from photographs (photogrammetry), videos (camera vector imaging), laser scans (light detection and ranging) and depth camera images (depth camera modelling) through on-the-ground and aerial data collection. On the consumer end, the mirror digital city can run professional applications that offer great value to the general public. An example is the use of avatar guides to offer virtual visitors city tours through attractions, stores, and restaurants. On the government end, virtual worlds serve as effective mirror world simulation tools for urban planning and adaptive policy planning.

There are certain drivers which can increase the adoption of virtual worlds in the future. Budget constraints continue to encourage enterprises to examine a wide range of alternatives to face-to-face meetings and public events. Coupled with travel restrictions and natural disasters, a growing opportunity has been created for hybrid events (events which combine a "live" in-person event with a "virtual" online component). In the short term, security and privacy issues pertaining to personal data are still being worked on and as such, enterprises are currently more comfortable with virtual world solutions hosted within their networks (also known as "private virtual worlds"). In the long term, interoperability across virtual worlds would help to speed up adoption as the avatars and digital identity developed in a virtual world can be re-used in another virtual world.

Biometric Authentication

With the need of employees to access data in corporate networks and perform financial transactions on the go, biometric authentication may grow in importance to strengthen the security of conventional authentication methods (e.g., password, one-time passwords). Biometric authentication may be categorised under two broad areas, namely physiological traits (fingerprint, iris, hand geometry) and behavioural features (voice).

Biometric authentication can provide medium to high levels of assurance but established, non-biometric alternatives (such as one-time passwords through a token) are available at a similar price point. Biometric authentication can free users from the need to remember passwords or carry a token. However, established physiological biometric technologies such as fingerprint and iris cannot be used reliably by every user due to poorly defined physiological features or alignment issues, necessitating the provision of alternative biometric methods based on behaviour (e.g., voice).

Voice biometrics is another authentication method premised upon the voiceprint generated by the individual's voice tract and cannot be exactly replicated even among twins. The individual is asked to repeat a randomly generated short phrase to generate a voiceprint, which will be used to authenticate the user through renditions of randomly generated short phrases.

Opus Research predicts the global number of voiceprints to more than double from 10 million today to 25 million by 2015. Voice biometric systems can leverage existing communication through telephones or mobile devices and can be automated using speech recognition systems. However, the main challenge is to reduce the false negative rates which are attributed to several reasons. These include voice samples of poor quality, the variability in voice due to illness, mood, changes over time, and background noise as the caller interacts with the system, and finally, changes in call technology (e.g., digital versus analogue). There is also the lack of established international standards. A standard application programming interface (API) is needed to reduce the issues with cost, interoperability, time-to-deployment, vendor lock-in, and other aspects of application development. Research institutes are working on addressing the long term challenges of voice biometrics, raising the performance level to a level comparable to other authentication mechanisms.

Wearable computers

Wearable computers and their interfaces are designed to be worn on the body to enable immersive user experience, mobility and hands-free activities. Traditional uses are for mobile industrial inspection, maintenance and the military. Emerging consumer uses include display peripherals for the consumption of digital content and the merging of the physical and digital worlds.

Vuzix has enabled users to watch videos in smart mobile devices through video eye glasses. More recently, it has developed the hardware and software for AR eye glasses, with a selling price of about US$5,000. MIT has also demonstrated SixthSense, a gesture-controlled necklace device that projects digital information onto real-world objects and locations. The main challenges for wearable computers are the lack of apps and services developed for these platforms and the relatively high price points which may deterw mainstream adoption.

Head-Mounted Displays

Head-mounted displays, as the name suggests, are designed to be worn on the head to enable an immersive user experience, mobility and hands-free activities. Traditional deployment lies in mobile industrial inspection, maintenance and military engagement. Emerging consumer uses include display peripherals for the consumption of digital content and the merging of the physical and digital worlds.

Vuzix has enabled users to watch videos in smart mobile devices through video eye glasses. More recently, it has developed the hardware and software for AR eye glasses, with a selling price of about US$5,000. MIT has also demonstrated SixthSense, a gesture-controlled necklace device that projects digital information onto real-world objects and locations. The main challenges for wearable computers are the lack of apps and services for these platforms and high price levels.

In May 2012, Intel Corp announced that its Intel Collaborative Research Institute for Computational Intelligence would be collaborating with specialists from the Technion in Haifa and the Hebrew University in Jerusalem, to conduct research in machine learning aimed at enabling new applications such as “small, wearable computers that can enhance daily life.”[21] According to Intel, such devices, which continually record what the user is doing, will be available by 2014 or 2015.

---

[21] Tova Cohen. Intel eyes future with computers that learn. [Online] Available from:

Google Glass, an AR glass (for the right eye) enabling users to manipulate and consume information through head gestures and voice input, is still in the prototype stage and could be ready for public release in 2014[22,23].

Brain computer interfaces

Brain computer interfaces interpret distinct brain patterns, shifts and signals —usually generated by voluntary user actions — as commands that can be used to guide a computer or other device. Non-invasive techniques use a cap or helmet to detect the signals through external electrodes.

In the USA, the University of Maryland is currently developing a “brain cap” technology that allows users to turn their thoughts into motion. The research team has created a sensor-lined cap with neural interface software that will soon be used to control computers, robotic prosthetic limbs, motorised wheelchairs and even digital avatars. Emerging techniques that measure electrical signals with no physical or resistive contact will drive advances in the longer term.

Though brain computer interface for elementary movement actions can be realised in three to five years, the use of brain computer interface to execute other commands may take a longer time. The major challenge for this non-invasive technology is obtaining a sufficient number of distinctly different brain patterns to perform a range of commands — typically, about five patterns can currently be distinguished.

The European Union’s eHealth R&D Group of Digital Barcelona, is working on an ICT-based solution controlled via a brain computer interface to create an assistive environment[24]. The prototype includes a brain computer interface interacting with: (1) home appliances (e.g., commercial television and a lamp) (2) a virtual avatar acting in a virtual representation of the user’s home and (3) today's modern Web 2.0 networks with the access to micro-blogging services such as Twitter.

The increased possibility of ICT being used to translate brainwaves into actionable outcomes through brain computer interfaces gives rise to several debates about the blurring of boundaries between human beings and technology. These debates could be related to the acceptable amount of technology in the bodies, the definition of human identity and the

---

http://www.reuters.com/article/2012/05/24/us-intel-israel-idUSBRE84N05520120524 [Accessed 9th July 2012].

22 Nick Bilton. Behind the Google Goggles, Virtual Reality. [Online] Available from: http://www.nytimes.com/2012/02/23/technology/google-glasses-will-be-powered-by-android.html?_r=4 [Accessed 9th July 2012].

23 CNET. Google Glass Explorer Edition. [Online] Available from: http://reviews.cnet.com/camcorders/google-glass-explorer-edition/4505-9340_7-35339166.html?ttag=fbwp [Accessed 9th July 2012].

24 BrainAble. *Autonomy and social inclusion through mixed reality Brain-COmputer Interfaces: Connecting the disabled to their physical and social world.* [Online] Available from: http://www.brainable.org/en/Pages/Home.aspx [Accessed 9th July 2012].

protection of brain pattern information from abuse and misuse. Before the brain computer interface moves into the mainstream consumer space, ethical and regulatory frameworks (e.g., addressing liability issues such as responsibility for unintended consequences resulting from "faulty" brain computer interfaces) would need to be developed. In addition, the governance mechanism to regulate the safety of brain computer interface devices would need to be established.

11.3.3 Five to 10 years

Advanced natural language question and answering

Natural language question and answering provides users with a means of asking a question in plain language of a computer or service and receiving an answer of high reliability and confidence in a reasonable time frame. IBM's iconic performance with its Watson supercomputer trumping the reigning champions on the television quiz show, "Jeopardy", captured the attention of the world. It was a critical milestone in the evolution toward effective cognitive reasoning by artificial actors. IBM had commercialised its Watson system through health insurer WellPoint (which has a 34 million client base), assisting medical professionals in diagnosing and treating patients. In February 2012, IBM indicated that "continuous human interaction" would be "the primary way these new computers will learn, enabling them to provide the information a user or company needs through experience rather than through scripts and commands." It is expected that Watson would be able to solve scenarios rather than tackle only specific questions in future and integrating with the cloud, would continue to learn even when passive.

Other emerging question and answering systems are EAGLi for the biomedical sector[25] and Wolfram Alpha[26], a general computational knowledge question and answering system.

---

25 EAGLi. Question-Answering Examples.

26 Wolfram Alpha. *Examples by Topic.* [Online] Available from: http://www.wolframalpha.com/examples/ [Accessed 9th July 2012].

Figure 9: IBM's Watson competing with Jeopardy champions[27]

Natural language question and answering systems are based on semantic technologies, which analyse and derive meaning from digital data through ontologies. These ontologies demonstrate relationships between entities in the data, making large amounts of appropriate digital information from various sources (e.g., social media, government data and real-time sensor data) easily comprehensible and contextualised to the ordinary human being. The long-term challenges and barriers to adoption will largely be related to issues of proof and trust. In particular, there is no broadly applicable methodology for making trust judgments when presented with information of varying quality, accuracy and confidence levels. There is also no common standard to represent information or bibliography sources and to manage the information flow it passes through the chain of custody from user to user. Furthermore, the challenges in effective interpretation of colloquial language and context, matching it to knowledge bases of potentially infinite scope, and the selection of a limited number of highly probable answers (even just one) remain profoundly difficult. In view of the challenges, natural language question and answering will not see mainstream adoption until about five to ten years.

Advanced question and answering systems can be used to boost the productivity of the customer-facing portions of organisations. In the midst of high call volume and multiple consumer requests, a substantial proportion of the requests could be handled by natural language question and answering systems, in turn freeing up organisational resources to perform value-added functions. Using voice and speech analytics together with these systems can help in the identification of difficult or dissatisfied customers for further remedial action by customer service representatives.

[27] IBM. *Watson, One Year Later.* [Online] Available from: http://www-03.ibm.com/press/us/en/presskit/27297.wss [Accessed 9th July 2012].

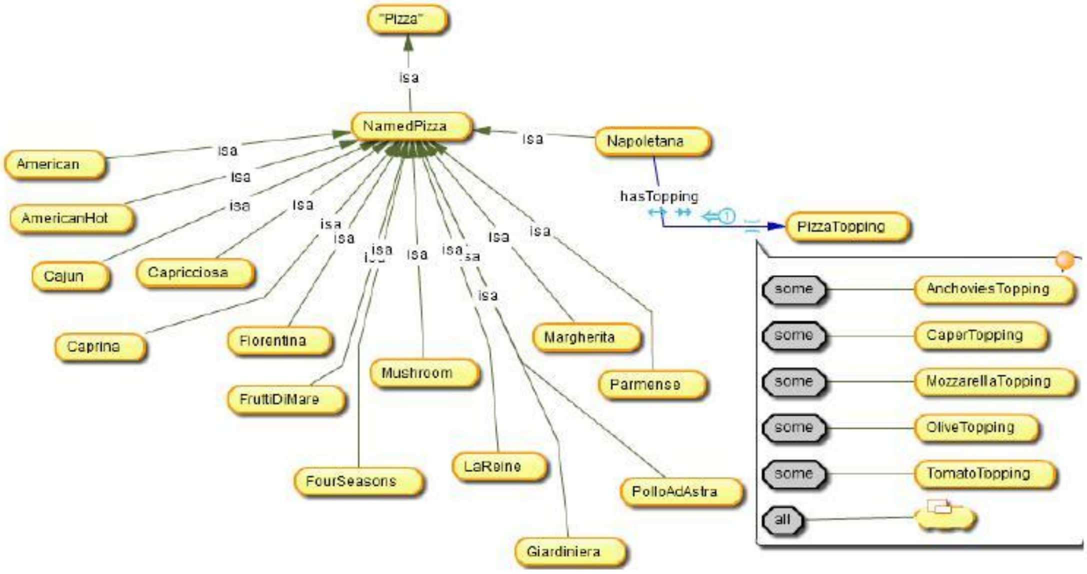

Figure 10: An example of an ontology in semantic technologies[28]

The convergence of virtual reality, natural language question and answer systems and computer vision could lead to the integration of virtual humans in virtual worlds. Interaction with virtual environments should now involve an immersive environment where users would interact with voice, gesture and physiological signals with virtual humans that would help them explore their digital data environment. Virtual humans should be able to recognise user gestures, speech and expressions and respond with speech and animation.

Immersive 3D telepresence

Immersive 3D telepresence provides a high-definition immersive experience, enabling the illusion of presence of geographically dispersed meeting participants in the meeting. The trend for telepresence is moving beyond business into social hubs (e.g., coffee clubs and dialogue cafes), retail malls (e.g., virtual receptionists), music collaboration, healthcare and education, which are all dependent on high quality video experiences. Although telepresence solutions in the market, such as Cisco TelePresence and Polycom RealPresence Experience, exist, there are opportunities to provide a more immersive experience by seamlessly integrating telepresence technology into everyday environments, providing 3D displays of meeting participants with full gaze contacts.

Immersive 3D telepresence may only be mainstream in possibly five to 10 years because of price points, connectivity (it requires dedicated high bandwidth connectivity) and quality of service demands.

The Sony Telepresence Room, at a price point of more than US$100,000, provides a fully immersive experience, with meeting participants appearing to be physically present and eye

28 Jon Modell. Ontology Visualisation: Tools for creating meaningful interactive diagrams. [Online] Available from: http://semtech2011.semanticweb.com/uploads/handouts/MON_445_Modell_3951.pdf [Accessed 9th July 2012].

contact with them aligned. In early 2012, Musion, the world leader in live 3D holographic projection, partnered with Masergy, a global provider of virtualised network services, to complete the world's first live 3D holographic telepresence using Polycom's high definition video technology. The immersive 3D telepresence experience was made possible through dedicated high bandwidth connections.

Ensemble interactions

With the proliferation of personal mobile devices, ensemble interactions will enable seamless user experiences across multiple personal devices. The vision of ensemble interactions is that interactions will easily shift to the most appropriate device as and when required by the user. This means that context changes experienced by users will drive ensemble interactions which may, in turn, require new sensors and control mechanisms to detect and automatically trigger the shift to alternative devices. The growth in smart IP-enabled TVs, and the increasing availability of wireless networking in computing and entertainment devices, suggest that home media will be an early adopter of ensemble interactions. Early examples that demonstrate this capability are gestural controls that flick a video from a mobile device to a large screen TV. The long term challenges pertain to undefined and uncertain user expectations about how multiple devices should collaborate in delivering an experience, and the limited number of mobile devices that can participate in such interactions.

Haptics

Haptics is the use of tactile interfaces (e.g., vibration or pressure) to provide touch or force feedback as part of the user interface. It may be adopted primarily in vertical-specific areas such as medical, surgical and dental fields, design, sculpting, computer-aided design (CAD)/ computer-aided manufacturing (CAM) and training/simulation.

In October 2010, the McGill University Health Centre (MUHC) announced that a completely robotic surgery and anesthesia procedure had been performed on a prostatectomy patient at the Montreal General Hospital in Canada. The DaVinci surgical robot enables surgeons to work from remote locations through tactile feedback while the anesthesia robot, nicknamed McSleepy, provides automated anesthesia.

In consumer mobile devices, haptics will enhance the user experience. For instance, Senseg Ltd., a start-up, is able to generate extremely refined textures through tactile feedback, creating a range of touch sensations from sticky to slippery and from glass-smooth to sandpaper-rough[29]. Senseg's haptic capabilities have yet to be implemented in consumer devices.

According to a 2012 market research report, "Touch Controller IC Market - Global Forecast & Analysis (2011 – 2016)", the total touch controller integrated circuit (IC) market, stimulated

---

[29] Robert L.Mitchell. *Display tech to watch this year: Haptics create a buzz.* [Online] Available from: http://www.computerworld.com/s/article/9213884/Display_tech_to_watch_this_year_Haptics_create_a_buzz?taxonomyId=128&pageNumber=1 [Accessed 9th July 2012].

by the emergence of natural user interface devices, is expected to reach US$7.7 billion by 2016, growing at a CAGR of 42.8% from 2011 to 2016. More than 80% of this revenue is contributed by the consumer electronics segment which has seen the rapid adoption of touch screens for mobile phone and tablets.

Figure 11: Da Vinci surgical robot[30]

Mass adoption of haptics in connected consumer electronics devices will be limited in the next five years. The key challenge is whether there will be a seamless inclusion of haptics in standard interface components (e.g., touch pads, mice or 3D gesture controllers) to control other consumer devices and to create touch sensations.

Speech-to-speech translation

Speech-to-speech translation involves translating one spoken language into another. It combines speech recognition, machine translation and text-to-speech capabilities. Approximate speech-to-speech translation has emerged through low-cost apps for mobile phones from companies such as Cellictica (Trippo VoiceMagix), Jibbigo and SpeechTrans. These are shown to be viable for consumer and tourist applications. In addition, Google has announced a "Conversation Mode" for its Google Translate Android app, and NTT Docomo has demonstrated cloud-based near real-time simultaneous translation of Japanese and English.

While there has been limited adoption of the technology by enterprises due to accuracy limitations, the availability of the low-cost mobile consumer products may drive interest in higher-end applications. As such, consumer mobile applications will likely be the first to attract significant interest. This will lead to higher end enterprise applications, including on-site interactions for field workers, as well as government security, emergency and social service interactions with the public. In the longer term, multinational call centres and

30 daVinciSurgery. *The da Vinci Surgical System.* [Online] Available from: http://www.davincisurgery.com/davinci-surgery/davinci-surgical-system/ [Accessed 9th July 2012].

internal communications in multinational corporations will benefit, particularly from routine interactions.

Natural language programming

The envisaged surge in tablet adoption and other smart mobile devices will leadusers to develop software programmes and apps directly on these devices. However, given the limited screen size, information layout and difficulties in typing on virtual keyboards, these devices are not optimal in enabling users to create new software programmes and apps productively. As such, there is a broader paradigm change to focus on programmers' needs and then build the tools to serve their needs, improving productivity and access to software and app development.

There have been several efforts in the past that incorporated speech interfaces for programming. At Canada's National Research Centre, a team of researchers developed VoiceCode, an open source software which allows programmers to write, edit and debug programming code through a voice interface[31]. Microsoft has also developed a spoken variant of the Java programming language called Spoken Java[32] and a plug-in named SPEED[33] for the Eclipse development environment to enable speech input. However, these attempts were based on programming languages that were not meant to be spoken, with punctuations that were difficult to represent by speech. To overcome the challenges of using speech for programming, a programming language can be created, using natural language phrases and an associated programming environment in which speech input is an instrumental part of the environment design .

Virtual laser keyboards

Virtual laser projectors are small projector modules that displays user interface (e.g., keyboard and mouse) onto any flat surface so that the users interact more easily with mobile devices or systems without the need for a physical keyboard or mouse. The laser projectors synchronise with mobile devices via Bluetooth and create a virtual keyboard almost the size of any standard physical keyboard on any surface. This gives users the freedom of space to type and the convenience of not hauling around a keyboard. Keystrokes on the virtual keypad projection send signals to the corresponding mobile device via Bluetooth. The keyboard can even have the functionality of a mouse which moves the cursor on the device as the fingertip moves on the laser surface. Existing products include Elecom's laser projection keyboard and Celluon's Magic Cube. Challenges to mainstream adoption include issues related to projection brightness under various lighting conditions, resolution

---

31 National Research Council Canada. *VoiceCode.* [Online] Available from: http://www.nrc-cnrc.gc.ca/eng/projects/iit/voice-programming.html [Accessed 9th July 2012].

32 Andrew Begel, Susan L. Graham. *Spoken Programs.* [Online] Available from: http://research.microsoft.com/apps/pubs/default.aspx?id=75117 [Accessed 9th July 2012].

33 Andrew Begel, Susan L. Graham. *An Assessment of a Speech-Based Programming Environment.* [Online] Available from: http://research.microsoft.com/apps/pubs/default.aspx?id=75115 [Accessed 9th July 2012].

of projected keyboards, power consumption in battery-powered devices, cost and whether mobile devices can incorporate the virtual laser projection capability.

Figure 12: Celluon's Magic Cube (Virtual laser keyboard projector)[34]

11.3.4 Ten years or more

Human-Robot Interfaces

The convergence of 3D location and contextualisation, natural language question and answering systems, speech-to-speech translation, gesture recognition, eye tracking and facial recognition may lead to the emergence of a truly intelligent human-robot interface. This could spawn a new generation of robots that are able to navigate in various physical environments, perform complex mechanical tasks (including taking care of the aged and the young) and communicate effectively with humans. With cloud computing providing scalability and Internet class computing, robots can access vast amounts of processing power and remotely gain new skills and behaviour. This approach will allow robots to offload computationally intensive tasks like image processing and voice recognition to the cloud and even download new skills instantly. By offloading heavy computing tasks to the cloud, hardware will be easier to maintain and Central Processing Unit (CPU) hardware upgrades will be hassle-free. Robots will have longer battery life and have less need for software pushes and updates.

Although cloud computing can enable cheaper, lighter and smarter robots, cloud robotics is no panacea. Tasks that require real-time execution still require onboard processing (e.g., controlling a robot's motion which relies heavily on sensors and feedback). Robots will still have to react to the physical environment with a significant amount of local computation.

[34] Celluon. *Magic Cube.* [Online] Available from: http://celluon.com/products.php [Accessed 9th July 2012].

We are witnessing some emerging international developments that use human-robotic interfaces for social interaction. Gostai, a French robotics firm, has built a cloud robotics infrastructure called GostaiNet which allows a robot to perform speech recognition, face detection and other tasks remotely. The small humanoid Nao by Aldebaran Robotics will use GostaiNet to improve its interactions with children as part of a research project in an Italian hospital. Readybot has their Collaborative Cloud Robotics technology, a software platform that allows human operators to use a video-game interface to control human-sized robots that do real-world tasks. These robots may be used for eldercare, janitorial tasks, inventory control and manufacturing revitalisation. Human supervisors control these robots remotely over the cloud as these robots follow pre-scripted routines to handle most activities.

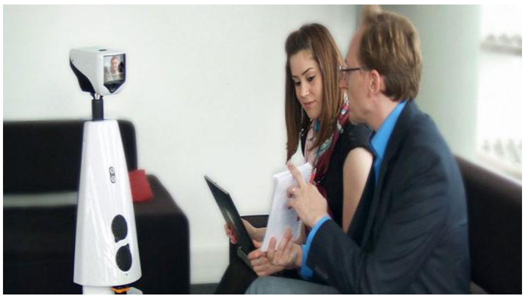

Figure 13: Robots leveraging Gostai's cloud robotics infrastructure[35]

The convergence of multiple infocomm enablers for new human-robotics interfaces can cause shifts in the current robotics landscape. Business and R&D communities now have access to modularised and interoperable intelligent algorithms, behaviour and skills to develop new human-robotic interfaces. This can level the playing field in robotics, speeding up the R&D process and lowering the time to market for new robotic products and services. From the enterprise perspective, local enterprises that seek to expand their manufacturing operations into the Asian regions can also operate robotic machinery remotely through the cloud, minimising travelling and maintenance costs. From the social perspective, robotics using cloud computing can now serve as an affordable alternative to meet household and healthcare needs. More of such robotic devices can have embedded sensors to measure health-related variables and using data and apps in the cloud, can aggregate and generate insights from various data sources.

Volumetric (beyond holographic) displays

Volumetric displays create visual 3D representations of objects, with an almost 360-degree spherical viewing angle in which the image changes as the viewer moves around. Sony

[35] Aldebaran Robotics. *Autism.* [Online] Available from: http://www.aldebaran-robotics.com/en/Projects/autism.html [Accessed 9th July 2012]; Gostai Jazz. *Remote Presence System.* [Online] Available from: http://www.gostai.com/connect/ [Accessed 9th July 2012].

demonstrated its RayModeler device but this, like most other volumetric displays, remains firmly in the lab environment. Holographic displays can create reasonable illusions of volumetric displays but technically, they are not true volumetric displays.

11.4 Market applications

Given the enabling technologies for user interface and future interface technologies, there are opportunities for application in various sectors. This section will cover five sectors: education, healthcare, homeland and Infocomm security, market intelligence and retail.

11.4.1 Education

In the education sector, user interface and interaction technologies can enrich the delivery of knowledge and enable students to be assessed in more natural ways.

In less than three years, immersive learning environments will start to see mainstream adoption. Immersive learning environments provide the simulation of realistic scenarios and environments to give students the opportunity to practise skills and interact with other learners. Toolwire, a US-based company has developed natural assessment tools through immersive learning environments with illustrated characters and settings, and engaging storylines. These were developed for a broad range of curriculum subjects including IT, healthcare, law, business, communications and education.

In the three to five year timeframe, formative learning assessment can help teachers customise lessons, based on feedback and assessment data on the student's learning journey. Such interactive learning systems may be in the form of apps developed on tablet platforms, drawing upon historical and real-time assessment data obtained during class and examinations. This will give teachers insights into the topics each student has difficulty with and enable them to assign relevant additional practice questions through the system. An example of such a system is eInstruction's Insight 360, an interactive formative learning system designed for mobile devices. By seamlessly integrating instruction and assessment, teachers are able to tailor lessons based on the review of student performance in lessons or across an entire year.

In the five to 10 year timeframe, affective computing may automatically tailor lessons to the emotional state of a user (using cameras to track eye gazes and facial expressions) and respond by changing a quiz or recommending digital content to fit the perceived difficulty of the learner. Some preliminary work has been done on affective computing in education. In the USA, the University of California, San Diego, developed a Computer Expression Recognition Toolbox (CERT), a fully automated facial expression recognition system that operates in real time. The system automatically detects faces in the video stream and codes each frame with respect to 40 continuous dimensions, including basic expressions of anger, disgust, fear, joy, sadness, surprise, contempt, a continuous measure of head pose, as well as 30 facial action units (AUs) from the Facial Action Coding System[36]. This system has been

[36] Littlewort, Wu, Whitehill, Fasel, Movellan, Bartlett. *The Computer Expression Recognition Toolbox.* [Online] Available from: http://mplab.ucsd.edu/~marni/Projects/CERT.htm [Accessed 9th July 2012].

used to evaluate the students' perceived difficulty levels with the curriculum through tutorials conducted by an automated tutoring system.

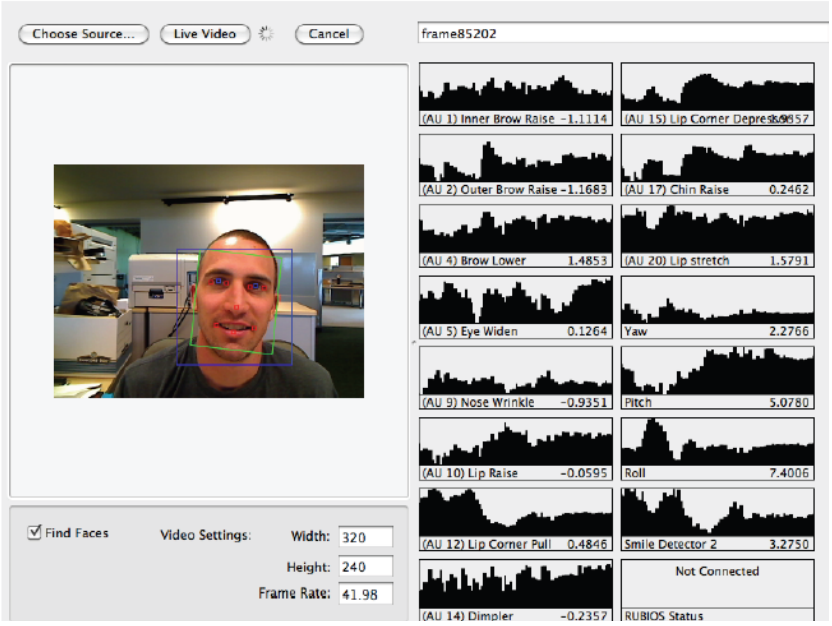

Figure 14: Facial recognition for detection of students' perceived difficulty levels in automated teaching systems[37]

11.4.2 Healthcare

In the healthcare sector, user interface and interaction technologies can help become the touchpoint for healthcare consumers and even help in the rehabilitation process.

In the three to five year timeframe, virtual medical assistants enable healthcare consumers to receive answers to their medical questions in a natural-language dialogue. The value proposition is the precision of answers and the ability to understand a broad range of linguistic phenomena. The assistants have a realistic 3D appearance, react continuously with emotional expressions appropriate to the conversation, have distinct personalities and speak with a natural voice. Virtual medical assistants have been deployed by the health service of Andalusia, a Spanish province with 8 million inhabitants. The Andalusian virtual medical assistant provides a range of medical, nutrition and childcare advice, helps healthcare consumers arrange, modify or cancel doctors' appointments, and deals with compliments and complaints.

Also, in the three to five year timeframe, brain computer interfaces can help in stroke rehabilitation and the process of motion recovery for stroke patients . The concept of this

37 Littlewort, Wu, Whitehill, Fasel, Movellan, Bartlett. *The Computer Expression Recognition Toolbox.* [Online] Available from: http://mplab.ucsd.edu/~marni/Projects/CERT.htm [Accessed 9th July 2012].

therapy is that when the patients imagine their movements, the same neurons used for movement execution will be recruited, speeding up their recovery.

11.4.3 Government

In government, user interface and interaction technologies can provide an environment for the seamless exchange of information and collaborative decision making across homeland security agencies.

In the three to five year timeframe, high-precision information environments can be used for command and control, and adaptive policy planning through the overlay and manipulation of relevant digital information over highly photo-realistic 3D environment models. With available technologies to map environments in high definition and 3D, a common perspective of the physical environment can be shared across agencies, enabling more collaborative and evidence-based decision making.

The fourDscape, developed by Balfour Technologies, is able to manage a large number of cameras and other sensors in a virtual, high-resolution, 3D display on a computer. The system provides a photo-realistic 3D environment overlay with additional data layers, such as street names and building locations. First responders using fourDscape can also monitor video cameras at the incident scene, participate in video conferences with colleagues, and set alerts to receive contextual and interactive updates. The updates can help incident commanders make tactical decisions and understand better what is happening on the ground.

11.4.4 Market intelligence

In the market intelligence sector, user interface and interaction technologies can help enterprises determine the effectiveness of consumer touchpoints and marketing campaigns.

In less than three years, eye tracking may be used in market research and website design to determine the effectiveness of alternate content and layouts. EyeTracking, Inc. completed a comprehensive evaluation of an American airline's e-mail communications, online advertising and self check-in airport kiosks through eye tracking[38]. The evaluation has led to improvements in feature visibility, visual attention and streamlined check-in processes.

Also, in less than three years, eye tracking and facial recognition may be used to profile consumers and recommend appropriate products and services. A gesture-based prototype application developed by the Fraunhofer Institute from Germany is a 3D camera system that identifies and then records the 3D positions of the hands, faces and eyes of target subjects. Two cameras record the face and eyes while the other two record hand motion. The interactive shop window is able to identify how many people are in front of the shop window, make suggestions of products on the basis of the profiled prospects and offer

[38] EyeTracking I. *Case Studies.* [Online] Available from: http://www.eyetracking.com/Case-Studies/Marketing-Research [Accessed 9th July 2012].

customised greeting texts on the display to attempt to engage and transform prospects into customers.

11.4.5 Retail

In the retail sector, user interface and interaction technologies enrich the shopping experience, with consumers having a better sense of what products can suit their needs with less hassle and effort on their part.

In three to five years, through virtual try-on shopping enabled by AR, consumers can intuitively search for clothes and virtually try them on to see if they fit. This is computationally demanding as the virtual try-on needs to take into consideration the consumers' and clothes' dimensions, how the clothes drape over the consumers' body contours and how the clothes move with the consumers (e.g., when walking and squatting).

At the same time, seamless indoor and outdoor location-based services are able to determine where consumers are and, together with other contextual data, provide more sophisticated value-added services. For example, a location-based service can detect that a consumer is on the first floor of a shopping mall and accessing the consumer's calendar, knows he has plans to visit another site. However, with the knowledge of the heavy downpour (real-time weather conditions) and the jam on the expressway (real-time transport conditions), the location-based service can suggest retail promotions happening hourly on a different floor in the mall as an alternative way to spend the time, given the potential delays on the road due to raffic and weather.

11.5 Implications

User interface and future interaction technologies are not new and we have witnessed a transformational change in how users interact with data, systems and services. These technologies create opportunities for improving Singapore's quality of life and enabling Singapore to remain economically competitive.

Consumers, can access information about real world objects, naturally and on-the-go, simply by pointing their smart devices at objects of interest. Emerging interaction technologies and the convergence with existing ones increase the options of how consumers interact with data and systems. Building mirror virtual worlds (highly photo-realistic models of the physical environment) with an overlay of digital information will yield useful insights for consumers.

For enterprises, the convergence of multiple user interaction technologies brings about richer data and information about the consumer and the operating environment. With increased contact points and channels, the market reach of enterprises is now extended and a seamless perspective of the consumer can be pieced together to generate insights for more focused, value-added services. With a sophisticated infocomm-savvy population and a racially diverse society, Singapore is well positioned to be a testbed for enterprises developing innovative user interfaces to subsequently penetrate he rest of Asia. It could also be a hub for talent, bringing together talent from multiple disciplines to develop a rich spectrum of value-added services.

For the government, user interface and interaction technologies can surface details of emerging phenomena on the ground which will otherwise not be apparent. This may be useful in providing a common operating view of the physical environment for collaborative command and control (e.g., events of infrastructure breakdown and homeland security incidents). User interface and interaction technologies can also enable adaptive policy planning by simulating the effects of policy and interventions on the existing environment.

In terms of the **enabling technologies**, there are already existing capabilities in the market. These include surface computers, interactive visualisation, voice and speech analytics, elementary eye tracking and facial recognition, all of which will continue to see increased adoption in less than three years. The enabling technologies in the three to five year timeframe will position the user more strongly as the new user interface and make sense of complex and large interconnected datasets for natural language question and answering systems. By monitoring and enhancing such capabilities in this timeframe, more advanced applications and services may be developed in sectors beyond gaming. In addition, inference capabilities on big data can enable intelligent question and answering systems which may be applied to a wide range of sectors.

Enabling technologies in the more than five year timeframe can be areas for R&D as these areas (e.g., emotion detection, voice biometrics) are not yet mature and can take at least five years to reach consumers' or enterprises' expectations in terms of either price, quality or both. Beyond the various enabling technologies, intersections with other disciplines (e.g., psychology, sociology and anthropology) need to be explored to better understand how consumers want to interact with data, systems and services. Singapore would need more professionals equipped with a system of systems thinking so as to encourage more sophisticated, value-added user interfaces to be developed for adoption across various sectors.

In terms of **market opportunities**, education, healthcare, government, market intelligence and retail are examples of sectors that can use user interface and interaction technologies for sector transformation. Some opportunities, such as high precision information environments for command and control, and adaptive policy planning, and seamless indoor and outdoor services, require coordination across sectors.

It is necessary to understand the underlying business processes, the required data for integration and stakeholder relationships. In the journey of understanding the current state and envisioning the future state, ontologies (known as shared representations of knowledge, showing entities and their relations) can be created to visualise the entities in the data and how they are related to one another.

For user interfaces and future interaction technologies to overcome **challenges** and encourage adoption, more collaboration is needed to determine how their ontologies and business processes are related to one another. Ontologies will need to be aligned and synchronised so as to form cross-sectoral ontologies that can be used across sectors. This will widen the perspective of sectoral transformation to also include an understanding of cross-sectoral linkages and their influence on one another in terms of operations and strategic economic impact.

On data use, policies on trusted data management and data governance will provide assurance that consumer data will not be used for purposes than otherwise intended. Technologies can support trusted data management and data governance such as data anonymisation to support downstream analysis, data fusion, and value-added services e.g., 3D location-based services. Furthermore, a secure, trusted and robust infocomm infrastructure has to be continually leveraged and improved. As many of such user interactions and services require intense computational resources and bandwidth, high quality of service of ultra-high speed broadband will enable services to be continually available to users.

R&D on user interface and interaction technologies can push the envelope on how users interact with data, systems and services. The promising areas for research institutes and universities are the enabling technologies in the three to five year and five to 5-10 year timeframes. Although there is no one single dominant user interface enabler or mobile platform, speech interface appears to be critical for the other enablers. Speech interface, although in nascent use today because of performance issues, will play a more important role in the future, e.g., in sophisticated and intelligent question and answering, speech-to-speech translation, software and service development, alternative authentication mechanisms and possibly even in lie detection.

Beyond the individual enabling technologies, the R&D community can work with the end-user community to better understand data needs, business processes and architect an R&D framework for the coherent orchestration of the various enabling technologies. In addition, the R&D community can examine the need for a platform in an artificial environment to aggregate the various enablers and showcase how these enablers may converge to solve a real-world problem. Such a platform can house modular and interoperable components for user interface technologies such as new algorithms for eye-tracking.