MULTIMODAL USER INTERACTION FRAMEWORK FOR CONTEXT AWARE E-COMMERCE

Supun Gothama Sirisena Hewawalpita
168063M

Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

January 2019

Declaration

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 16-01-2019

Name: S.G.S. Hewawalpita

The above candidate has carried out research for the MPhil Dissertation under my supervision.

Signature of the supervisor:…….. Date: 16-01-2019

Name of the supervisor: Dr. G. I. U. S. Perera

Abstract

E-commerce has grown up to be a major use of e-services and online purchases through the e-commerce are largely preferred over the traditional brick and mortar purchasing. Yet it is challenging for the consumers to fully experience the products or services with limited senses, lack of tangibility and sense of presence. Therefore a vital research question can be identified; how multimodal interactions can be used in e-commerce with context awareness, to improve the consumer experience.

To address that question, this research aimed to study multimodal interactions, contextual factors and their effects on consumers. A set of multimodal interactions including 3D visualization and hand gestures and related contextual factors such as user, access device were identified in this research. They have been used to develop a multimodal interactions enabled prototype e-commerce framework.

Several experiments and user studies have been conducted using the developed e-commerce framework and interesting effects on consumers have been discovered including positive user experience, improved value perceptions, and positive product opinions. Most importantly it has been shown that consumers perceive about 50% increased product value, and they are more likely to purchase when interacted multimodally. Usability Evaluations on the framework showed that users are mostly successful and comfortable in using multimodal interactions. Some technical, social and cultural barriers and challenges for enabling multimodal interactions were also revealed in those evaluations.

From the findings of this research, it is suggested that further research focus should be on overcoming the identified technical, social and cultural barriers and bringing multimodal interactions to mass usage in electronic commerce platforms. Also the multimodal interactive e-commerce framework developed in this research can be used as platform to further study consumer dynamics by changing various variables.

Keywords: Consumer experience, Context awareness, E-commerce, Multimodal interactions

Acknowledgments

I would like to express my sincere gratitude to my supervisors and mentor Dr. Indika Perera, for the continuous support given throughout this research. This would not have been a success without your incredible mentorship and advice from the beginning.

I would like to thank all the staff from the Department of Computer Science and Engineering for their help and guides provided on all occasions.

This research was supported by University of Moratuwa Senate Research Grant, and I sincerely thank for the financial support provided by it.

I am also grateful for all the participants in my research experiments and user studies, from the University of Moratuwa as well as from outside, who spent their valuable time for this research.

Special thanks to my family: my parents, brothers, and wife. Your encouragement and understanding were enormous from the beginning. I would also like to thank all my friends who made my time in research lab wonderful in the past two years.

Table of Content

List of Tables

List of Figures

List of Abbreviations

3D	Three Dimensional
API	Application Programming Interface
AR	Augmented Reality
B2C	Business-to-Consumer
CMS	Content Management System
DNN	Deep Neural Networks
DOF	Degree of Freedom
HCI	Human Computer Interactions
HMD	Head Mounted Display
HMM	Hidden Markov Models
MOOC	Massive Open Online Course
RFID	Radio-frequency identification
VR	Virtual Reality
XML	Extensible Markup Language

1. INTRODUCTION

1.1 Background

In the past decade, e-services have grown up to be a major application of information and communication technology with the widespread availability of the internet and have become an essential part of our lives with the rapid advances in telecommunication technologies and the Internet. Many things we do day to day, knowing or unknowingly involve some kind of an e-service, be it an e-commerce, e-channeling, ticket booking or an internal service of a business process.

Out of all these services, e-commerce dominates as the most widely used e-service. According to statistics, half of the global internet users have made an online purchase in 2016 [1]. E-commerce is an online service where transactions of buying and selling are possible, unlike traditional websites which only have descriptive information [2]. Although e-commerce reminds us selling of tangible goods online, a large proportion of e-commerce is about consumption of services and experiences.

Unlike traditional shopping, e-commerce services are not constrained by the distance or time. Thus it gives more control over to the consumer. On the other hand, e-commerce services are less tangible than the physical in-store shopping. As a result, the consumer cannot use all the senses to experience the service or good as in a traditional shopping, rather has to rely on sight and sound. It is challenging for the consumers as well as for the designers of e-commerce services. In the end, the quality of e-commerce services depends on how the users perceive and experience the service as well as usability and other HCI (Human-Computer Interaction) aspects. Because when there is a human user involved, his perception about the service and interactions with it affect the final outcome [3].

1.2 Motivation for the Research

Human interaction with the world is inherently multimodal. That is, we use multiple senses to explore and perceive our environment. But human-computer interactions

have historically been unimodal and primarily through a single mode such as text on a screen with a keyboard for input. Lately, systems are rapidly adopting multimodal interfaces to enable free and natural interaction using the natural human capabilities of communication [4]. Having multiple modes of interaction in a system gives more affordance and greater accessibility to the users and brings more bandwidth to the communication.

Unlike stationary computers of early days of computing, in modern days users take their mobile devices with them. Thus the operating environment and situations frequently change. Hence systems had to adapt to provide same functionality independent of the environment or improved functionality depending on the environment. Context awareness allows users to interact with systems which are aware of the environmental states such as devices, location, and services.

Context awareness can be used for combining different modalities of input and output and for interpreting them. Together these are emerging as major trends in the realization of ubiquitous computing systems.

1.3 Research Statement

There have been many researches on using multimodal interactions and context awareness to address various issues of e-services. But rather than improving the quality and user experience of an overall services process they focus on addressing an individual interaction or part of a service process. Typically e-services are complex business processes and improvements to isolated user interactions do not achieve the desired user experience. To achieve this overall quality and user satisfaction, multimodal interactions and context awareness have to be introduced, considering the whole e-service and its business process and also the business process has to adapt accordingly.

Developing or updating an e-commerce platform with such interactions and context awareness features can be a complex and tedious task. Specifically designed

framework with reusable components and guidelines can simplify the task to some extent, but there exists no such framework.

Hence the core research questions addressed in this study are,

1. *How multimodal interactions can be integrated into e-commerce with context awareness, to improve the consumer experience?*
2. *How to simplify the development process of such e-commerce systems by developing a customizable multimodal interaction framework?*

1.4 Research Methodology

First, we analyzed the importance of multimodality in e-services and its impact on the overall quality of services. We then looked at related research work conducted in the same area, their results, and findings. Then we identified the possible modes of interactions and evaluate their practicality, advantages and drawbacks under different scenarios of use. User evaluations and feasibility studies are conducted to identify some of these factors. Accessibility factor were given a special consideration wherever possible in this research, so interaction modes for users who experience disabilities are specially considered. Then we identified and defined suitable contextual factors and how they affect the mode of interactions and service outcomes.

Based on the results of the previous steps an interaction framework was modeled and implemented in a prototype level. This framework consists of a set of tools, components, and guidelines that can be reused to implement multimodal interactive e-services. Then an evaluation scheme was modeled and the performance of the prototype was evaluated under a selected use case.

1.5 Objective of the Research

Most of the literature studies discussed in a later section had focused on individual interaction modes or context awareness features and using them in e-commerce scenarios. A gap in literature was identified about combining the existing interaction

modes into a complete e-commerce solution or a framework that can be readily used to deploy e-commerce websites.

The goal of this research was to examine how the overall quality of e-commerce could be improved with multimodal interactions and context awareness and develop a framework to simplify the implementation process. Rather than focusing on isolated system-user interactions, the idea was to improve the quality of a whole service process, even removing or introducing new interactions wherever necessary.

There were five major objectives of this research.

1. Examine the importance of multimodality in e-commerce
2. Identifying the suitable modes of interactions
3. Identifying the suitable context factors
4. Design and develop a multimodal framework for context aware e-commerce
5. Evaluate the framework for realistic scenarios and users

The successful completion of this project was expected to identify suitable multimodal interactions for e-commerce and context awareness factors that integrate them and evaluate their effect on users. Therefore an expected outcome of this project was to improve the consumer experience in e-commerce.

As another outcome, a multimodal interaction framework was developed to aid rapid development of multimodal interactive e-commerce applications. This framework consisted of boilerplate software framework and a set of guidelines on using multimodality. Also, this framework will aid any future studies on e-commerce and multimodality by facilitating a platform for researchers to study the effects by changing various interaction modalities.

1.6 Contributions and Articles

A multimodal interactions enabled e-commerce framework has been developed as a part of this research that can be used to rapidly set up e-commerce systems with multimodal interactions without developing from scratch. This framework sets the

ground for further research in fusing multimodality with e-commerce, as a tool for future experiments.

Further, the following articles have been published from the research so far.

- "Effective Learning Content Offering in MOOCs with Virtual Reality-An Exploratory Study on Learner Experience", S Hewawalpita, S Herath, I Perera, D Meedeniya. Published in Journal of Universal Computer Science 24 (2)
- "Effect of 3D product presentation on consumer preference in e-commerce", S Hewawalpita, I Perera. Presented at Engineering Research Conference (MERCon), 2017

1.7 Organization of the Thesis

The rest of the thesis is organized as follows. Chapter 2 reviews related work in the areas of e-commerce frameworks, multimodal interactions, context awareness and their usage in the context of e-commerce and other related e-services.

Chapter 3 explains the importance and the practicability of multimodality in e-commerce and their effects on the consumer. Selection of interaction modes and contextual factors to be supported by the framework is also described in this chapter with the preliminary studies conducted, and how they have influenced some of the design decisions about the framework.

Chapter 4 explains the implementation details of the multimodal e-commerce framework in detail with the software architectural views, how each interaction mode is implemented in the framework and the hardware devices that are supported by the framework. It also discusses how the context awareness is used to integrate multiple modes of interactions.

Chapter 5 discusses the evaluation scheme for the framework and present different experiments and user studies conducted with prototypes systems created with the framework.

Chapter 6 analyzes the results and data gathered from these experiments as well as presents the major observations and findings of the research. Chapter 7 summarizes the finding of this thesis and concludes with a look into future work and directions.

2. Literature Review

This chapter details the literature related to multimodality and context awareness in e-commerce and similar e-services and development of e-commerce frameworks. The chapter is divided into four main sections.

Section 2.1 introduces the limitations of existing unimodal e-commerce systems in providing consumer experience and how multimodality can be used to address this issue. This section discusses different modes of interactions, their theoretical backgrounds and the effects they have on the users.

Section 2.2 discusses the importance of accessibility in e-commerce and other widely used e-services and also about the challenges and issues in enabling such systems for differently abled users.

Section 2.3 delves into context awareness and how it has been used in the domain of e-commerce to enhance user experience. This section also discusses the importance of context awareness in integrating multiple interaction modalities.

Finally, section 2.4 explores e-commerce frameworks in literature, their architecture and key design decisions that have to be considered when implementing a framework.

2.1 Multimodal Interactions

One main problem of e-commerce is, the delivered services or goods are not very tangible as the same services or goods delivered physically in in-store shopping. The users have to experience the service through a limited number of senses and unable to feel, touch or manipulate through the web interfaces. It is an indirect experience provided through symbolic representations and mediated interactions using textual and graphical descriptions. In contrast, a direct experience should provide unmediated interaction between the consumer and product or service allowing using his full sensory capacity [5]. Indirect experiences are in most cases less effective than

direct experiences on the consumer [6]. Multimodality can be used to address this primary issue by expanding the modes of communications and the consumer experience.

In a simple sense, the term "multimodal" refers to the use of textual, audio and visual modes in a medium to deliver a message [7]. Collection of these modes directly affects how the receiver perceives the idea or concept. For an example, in the case of a music video, music, vocals, and moving images are different modes used to express an idea to the viewers and the combination creates new forms of literacy [8]. The association between words and symbols or the meanings known as is the fundamental of a language [9]. A multimodal message changes its semiotics effect by adding meanings in multiple modes or contexts, thus creating a different meaning to the receiver [10], [11].

Mutual disambiguation or the ability to use information provided by several modes to resolve ambiguities in each other is also a benefit, which thereby reduces errors of a system [12]. Thus multimodal systems are well suited to manage uncertainty of users and sensors.

There are number of interaction modes that are researched and experimented in literature. Some of the interaction modes have found their way into the mainstream consumer applications while some remains limited to the research contexts. The following sections summarize those researches under each interaction category.

2.1.1 Speech

Speech as a mode of communication for e-services has been researched for a while. Speech is probably the most natural communication mode for humans and allows the hands and eyes to be busy in a separate task. But in technology wise speech recognition is a challenging task, especially in noisy environments. Earlier speech-based systems were developed using VoiceXML technology which is a document standard for specifying interactive voice dialogs. Rouillard developed a prototype e-commerce system combining HTML and VoiceXML technologies [13]. It allowed

the users to interact with the website in traditional way while interacting with the voice service over the phone. Information that is difficult to convey vocally such as photos is displayed on the website synchronously. Another similar technology known as XHTML+Voice or X+V later became a standard which is an XML language for describing visual and auditory multimodal user interfaces. This is achieved by combining of JavaScript and XML Events.

Another research has combined VoiceXML with stylus input to develop multimodal voice/ink mobile e-service application [14]. The stylus is used to write words that are corresponding to speech or to convey symbols and signs. Partial information of each mode is used to correct errors is other modes and interpret the combined information. With the rapid growth of smartphones, voice interactions started its way into the widespread usage. Live Search for Mobile [15] was one of the first commercially available multimodal voice interactive search applications. They found out that users tend to use the application in noisy environments and typically in arm's length making speech recognition a difficult task. In an experiment scenario, application correctly identified entire sentences in only about two-thirds of the cases.

Although been widely researched throughout the last decade or so, automatic speech recognition is still the technically most challenging aspect when it comes to enabling speech interactions. Essentially it is a problem of recognizing the words spoken by a person, relying only on the information contained in the voice signal and on prior knowledge on the problem domain and may be a dictionary of potential words.

Speech recognition is commonly formulated as a statistical classification problem where a parametric representation of the speech audio signal is used to classify signals into allowable words. Hidden Markov models (HMM) are the mostly used parametric model at the acoustic level. Although Hidden Markov models are effective in acoustic results in good recognition results under many conditions, they also have their limitations as well [16].

Recently deep neural networks have been increasingly used in speech recognition and most of the modern state of the art speech recognition systems contain some form of deep neural networks [17]. Deep neural networks with many hidden layers trained using new methods have been shown to outperform a variety of traditional speech recognition benchmarks, even by a large margin.

A deep neural network is a feed-forward, neural network that has a number of layers with hidden units between its inputs and its outputs. Each hidden unit typically uses a logistic function to map its input from the layer below, to the output that it sends to the layer above. Training DNNs for automatic speech recognition takes considerable time and are trained by discriminative back-propagation algorithms. Recently researchers have shown that significant gains speech recognition can be achieved by adding an initial stage of generative pretraining to DNNs by reducing overfitting, and the time taken for fine tuning [18].

Now speech recognition has become inbuilt in most of the mobile and desktop operating systems and with the availability of APIs (Application programming interfaces), developing multimodal voice interactive services is more probable than before.

Meng et al. [19] have researched a new technique of using crowd improved speech recognition for e-commerce to support elderly people. They have used Web Speech [20] capability of modern web browsers with Microsoft Speech Platform [21] to recognize spoken queries in real-time. In the same research, they manage to provide personalized speech feedback with a synthesized voice.

Web Speech API [20] specification published in 2012 by Speech API Community Group has defined a JavaScript API to enable speech recognition and voice synthesis on web browsers. It provides developers a common API to generate text-to-speech output and to use speech recognition as an input in web applications. They have also provided security and privacy guidelines as well as implementation considerations.

Although it has been some years since the specification has been published, only a few major web browsers support the Web Speech API to date.

2.1.2 Gesture Interactions

Gestures have long been experimented and researched as a more natural interaction technique for humans to communicate with computers. In humans, hands movements and gestures play a vital role in interacting with the world as well as in communicating with others. The two degrees of freedom (DOFs) users get from a typical mouse cannot properly emulate this three-dimensional space.

The term Gesture Recognition generally means the whole process of identifying and tracking the human gestures made with hands or heads and converting that information to meaningful input commands for a program. Implementation of such a gesture recognition system is accomplished through two enabling techniques [22].

1. Contact-based devices
2. Vision-based devices

Contact-based devices rely on physical interactions by a user with a device which is typically worn by the user such as gloves. Also, there are devices that are held by the user such as gaming console remotes with accelerometers.

Myo armband [23] is one recent contact based gesture recognition device found its way into the consumer market. It is a band that is worn on the user's forearm and it uses eight electromyographic (EMG) sensors to measure electrical activity in the forearm muscles as well as gyroscope and accelerometers to detect gestures.

On the other hand, vision-based gesture recognition relies on video data and image processing to identify the motions. Vision-based gesture recognition is a challenging task [22] with many degrees of freedom, a variation of gestures, different lighting conditions etc. Also, it has to recognize gestures performed by humans of different sizes and colors with different variations. Most of the vision-based hand gesture recognition techniques adopt three phases: detection, tracking, and recognition.

Not all the vision-based gesture recognition systems rely on visible light; rather some uses infrared lights and sensors as well. Microsoft Kinect [24] sensors have inbuilt infrared emitters and cameras that capture visible light as well as infrared light and is able to detect the depth to the users in the environment. Leap Motion [25] is another such recent device relies on infrared vision. It can recognize and track hands and fingers and various gestures with sub-millimeter accuracy using its two monochromatic IR cameras and three infrared emitters. Whereas the Kinect is suitable for full body tracking in a larger environment, Leap Motion is suitable for more delicate hand and finger-based interactions with computers as well as in virtual reality environments.

Although gesture interaction designing is generally focused on taking user's gestures as an input modality, it is also been used as an output modality as well, especially with virtual agents with human-like avatars. Experiments have found that gestures made by humanoid virtual agents complement their human-like appearance and they promote friendliness, politeness, and lifelikeness [26].

2.1.3 3D Visualization

3D visualization is one technique that can be used when there are physical objects or products involved. This is achieved by using a number of photos from different angles or a 3D model of the product/object. Allowing viewing all parts of the object interactively have an effect on fast reasoning and understanding of the object by consumer. It has been found that in e-commerce services, 3D interactive product presentation generates a more positive attitude toward products [27].

Particularly the 360° views using the actual photos are found to be more appealing to the consumers than 3D models [28]. One possible reason for this might be the fact that as humans we tend to understand things by creating conceptual models in the mind. When the whole product is viewed from all the angles it helps fast reasoning of the mind. Also, Consumers interacting with such products are more likely to experience an increased sense of presence [29], [30]. These kinds of visualizations

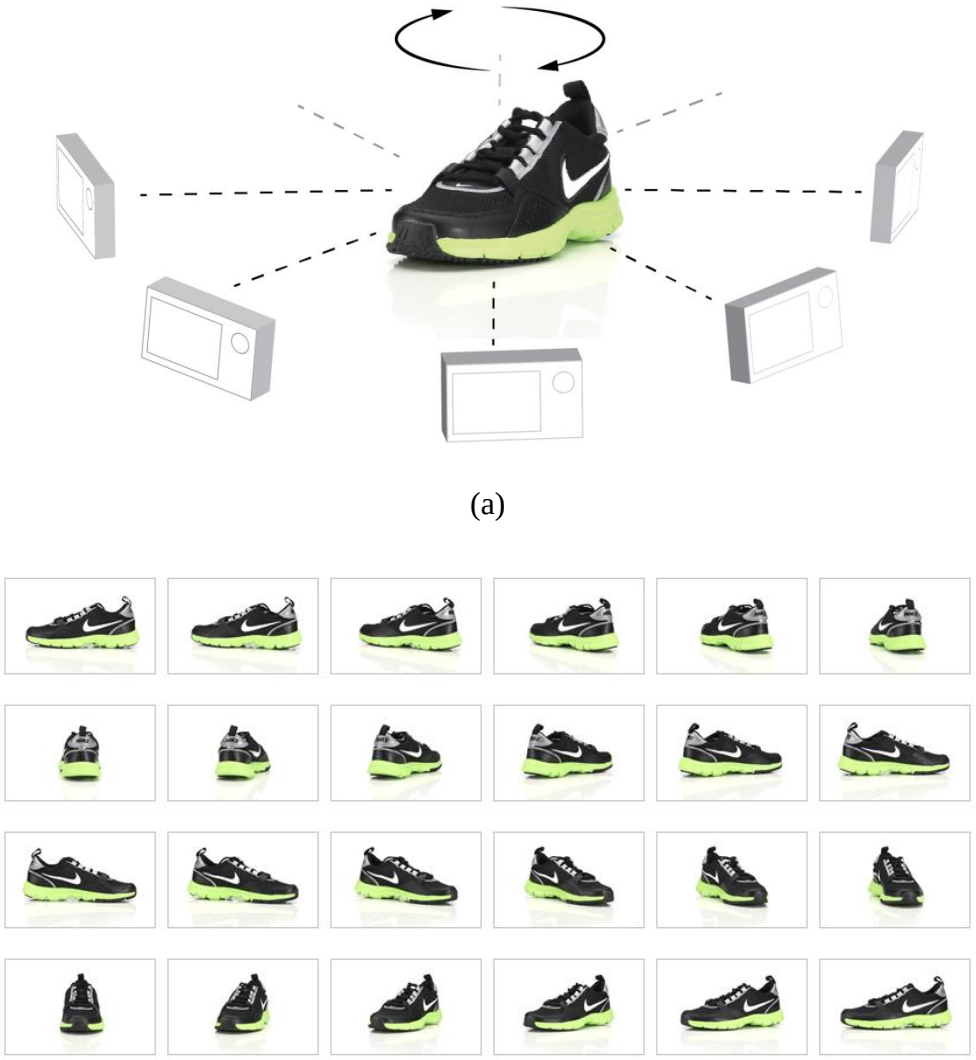
are emotionally stimulating and are especially suitable for promoting specific products. Positive emotions in users lead to an improved perception of the usability of the product and purchase intentions.

The basic principle behind the 360° view is mapping the 360 degrees of space to a set of actual images taken around an object at evenly spaced angles, thus creating a pseudo-3D visualization (see Figure 1). There is also complex variation in which the view can be rotated around the vertical axis as well the horizontal axis. Creating such visualization for a product is a laborious process and requires taking dozens of photos around the object in precision angles.

Many different software and hardware solutions are available to automate this process. The hardware solutions typically have a turntable stage where the product is positioned and with the aid of software, the stage rotated and photos are captured in even angles and automatically processed minimizing the manual effort and time required.

Some researchers have investigated the effects of interactive image technology (IIT) and suggest that higher level of IIT such as virtual 3D models are suitable for online apparel retailers to improve consumer perception and shopping experience [31], [32].

Providing effective user interfaces for 3d interactions is challenging in its own right. Zooming, rotating the view and manipulating the objects require feel-natural interaction techniques in order not jeopardize the user experience. Capabilities of the human motor system and cognitive conceptualization need to be considered when designing such interfaces. Khan and Nikov have proposed a mixed reality based customer interface for e-commerce, where users can interact and manipulate virtual objects by moving real-world objects [33]. They suggest using ‘hold in hand’ metaphor which mimics holding an object on hand and rotating it.



(a)

(b)

Figure 1: (a) Capturing photos around a product (b) Sample set of captured 24 images

2.1.4 Augmented Reality

Another related area of 3D based merchandise is the augmentation of products with 3D product models through augmented reality. The concept of Augmented Reality can be explained as a superimposition of computer-generated two dimensional or three-dimensional objects over the real-time scene acquired into a capturing device.

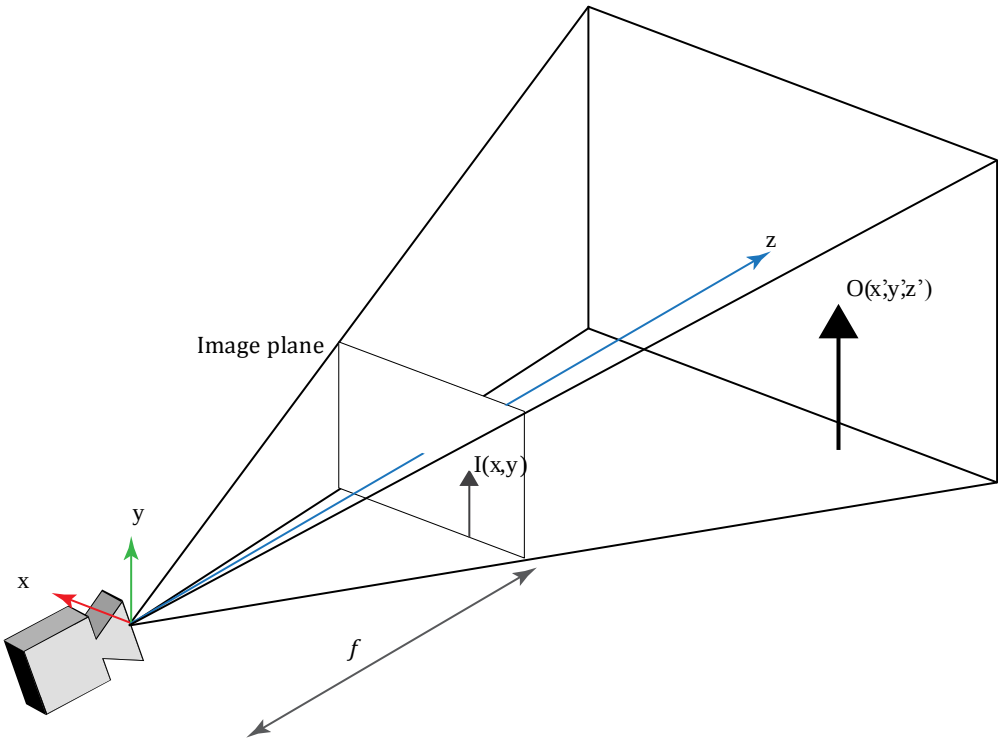


Figure 2 Simple pinhole camera model

The basic task of augmented reality is to position a camera in a virtual environment at the exact position of the viewer's eyes looking at the same direction (Figure 2). For that, we have to first determine the intrinsic and extrinsic parameters of the camera. The intrinsic parameters of a camera describe how the camera will convert objects within the camera's field of view into an image and encompass focal length, image sensor format and principal point. The extrinsic parameters describe the position and orientation of the camera in 3d space. To find the camera parameters a pinhole camera model can be used approximating the camera that produces a given photograph and it is known as camera calibration.

A key concern in developing AR systems is display technologies. There are hardly any see-through displays that have sufficient brightness, resolution, field of view, and contrast to seamlessly blend a wide range of real and virtual imagery. Furthermore, many technologies that begin to approach these goals are not yet sufficiently small, lightweight, and low-cost making them unsuitable for wide use. On the other hand challenges of video see-through display design is to ensure that the

user's eyes and the cameras effectively share the same optical path, eliminating parallax errors that can affect the performance of close-range tasks. Augmented reality to display 3D product information is mostly used in product stores with wearable devices or through the smartphone. These are done as either marker based, location-based or through feature identification [34].

Real-time fiducial marker tracking is a standard mechanism in augmented reality which relies on the mobile phone camera to detect markers of predefined format. A marker is usually a black and white grid patterns surrounded by a black border which correspond to a bit pattern. By identifying and calculating the position, angle and the id of the markers, 3D images are rendered on the screen to match the positioning and rotation of the markers giving the augmented reality effect.

Unlike natural feature tracking, detecting and decoding artificial markers [35] are highly robust and works well under varying lighting conditions and with minimal computational resources. Moreover, fiduciary markers' unusual visual appearance makes them more noticeable, helping users identify information hot spots in visually cluttered environments as well as large environments where interest points are sparse.

Feature-Based Augmented Reality relies on detecting naturally occurring features, with prior unknown positions to display augmented content. This process consists of three major steps [36];

1) Feature detection
2) Feature extraction
3) Matching descriptors

First, key-points on the camera feed are detected. Then a descriptor is computed for each key-point using the pixels around that point. Then the corresponding points between the two images are identified to calculate the position and angles. This process computationally intensive compared to the marker-based approach and less robust. Lighting conditions and occlusions make the same feature appear differently.

In Location Based Augmented Reality approach, augmentation of the rendered images depend on the three-dimensional location and orientation of the viewing device. The exact position of the device in the environment is estimated using GPS signals or indoor positioning technologies such as Wi-Fi or Bluetooth beacons. The orientation of the device can be calculated using accelerometer and magnetometer sensors inbuilt in modern smartphones and tablet devices.

Due to the low fidelity in the estimated positions, location-based augmented reality is not suitable augmenting imagery in close range. Rather it used to deliver augmented reality experiences within a large environment where the augmented objects are in a sufficient distance to neglect the positioning errors. AugLAC [34] is one such framework where they have used indoor positioning based augmented reality in a shopping complex environment.

There have been some efforts on fusing e-commerce with in-store shopping using augmented reality visualizations [34]. Augmented reality can introduce several advantages for physical store based product exploration but for e-commerce, it should be carefully designed to avoid redundancies from web-based information presentation and augmented 3D models.

Augmented reality can be used as an experience that generates value for consumers and increases consumer engagement. Scholz and Smith [37] have come up with a framework that describes the ingredients of augmented reality and basic design decisions that have to be considered when designing such experiences. Zhu, Wei, et al [38] have developed an in-store e-commerce system that uses augmented reality to provide shopping assistance and customized advertising. They have also used a context awareness technique which is an extension of the conventional advertising concept of changing the content based on the environment and interactions of users. This system, PromoPad, uses augmented reality on a hand-held Tablet to dynamically modify the context of products on store shelves by see-through vision with augmentations.

2.1.5 Virtual Reality

Virtual reality is a fast improving technology, where users immerse in a computer-simulated environment. The immersive environment that is generated can be a realistic representation of the world or can be imaginary, allowing the user to have an experience that is not possible in the physical real world. The believability of the virtual experience or the sense of presence is a key factor in VR. Currently, VR technology uses head mounted virtual reality headsets with a display for the eyes and gyroscope, accelerometer sensors to track head position and orientation [39].

The basic concept of virtual reality is providing a stereoscopic display to the eyes using a head mounted display (HMD). The displays are usually LCDs, one per each eye or in some devices one display with screen split for two eyes. Lenses attached in front of the display focus the visualization to the eyes in a way similar to how human eyes view the real world. Due to the stereoscopic effect of the display, two eyes get slightly different views of the virtual 3D environment, creating the perception of depth. And for the resulting perception to be realistic, a frame rate of at least 60 frames per second is required.

The video for the display is either generated on a computer or console and sent to the HMD through a cable or generated in the HMD itself. Commercially available high-end devices such as Oculus Rift [40] and HTC Vive [41] use the first approach and thus require a powerful graphics capable computer for rendering. On the other hand, lower end devices such as Google Daydream [42] does not require external graphics processing and rather they rely on an inbuilt processing unit or a smartphone placed in the HMD.

Providing a stereoscopic view is not enough to achieve an immersive experience in VR. To be fully immersive, the view has to change according to where the user is looking. For that head movements and orientations have to be tracked. Rotations of the head also known as pitch, yaw, and roll are tracked using sensors attached to the headset, such as accelerometers, gyroscopes, and magnetometers.

Unlike orientation, tracking the position of the head is not a straightforward task and most of the time requires an external reference frame to measure the relative movements. One simple mechanical approach is to connect the head mounted display with a fixed anchor using an articulated series of arms, where they fold and extend according to the user's head movements [43]. Although this approach results in very accurate results, the freedom and the comfort for the user are limited.

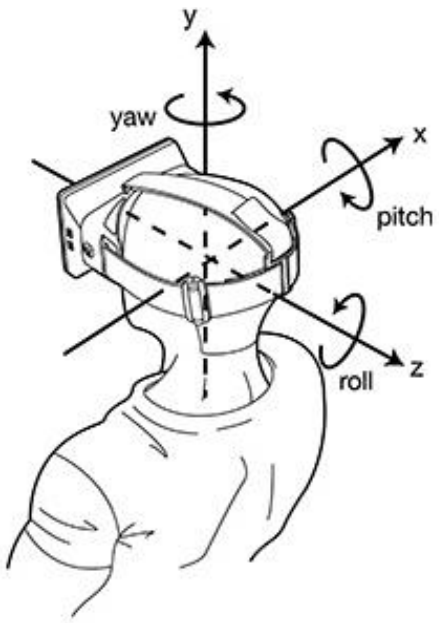


Figure 3 Six degrees of freedom in virtual reality

Most of the commercial virtual reality headsets rely on optical sensing to track head position. These tracking systems have two components: a light source (usually infrared) and an optical sensor. By sensing the reflected light off the user, the depth and angle to the user from the tracker are approximated.

Head-tracking techniques have to be low latency in the range of milliseconds to be effective or otherwise, the lag between the head motion and the virtual environment changes will be noticeable [44]. This lag typically arises due to transport delays, measurement processes of the various sensors and time taken for graphics rendering. Delays in virtual reality have significant impacts on user experience and performance. Although with high performing computers and graphics processing

units, the latency can be reduced for an acceptable level, the combined lag is impossible to eliminate completely. Thus some guidelines are used to determine an acceptable level of lag [43].

- Users feel the presence in the virtual world
- Fixed objects in the virtual world appear stationary while head movements
- No occurrences of motion sickness
- Task performance is not affected

If the above criteria are met, the lag can be presumed to be within an acceptable level and would not have a significant impact on users.

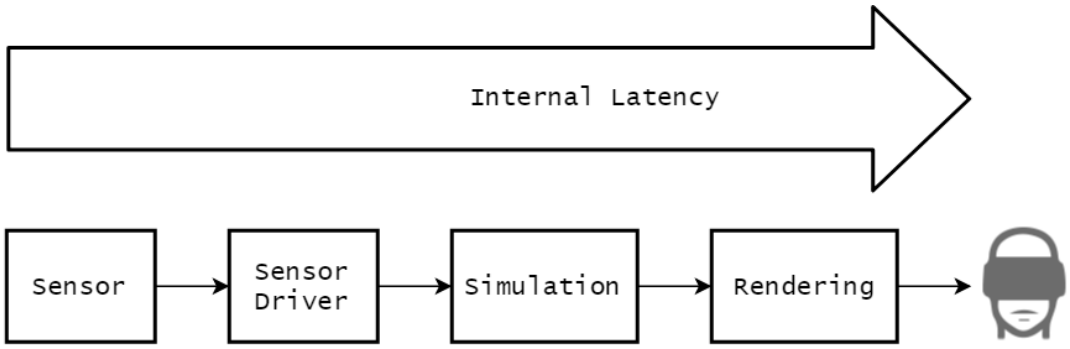


Figure 4 Lag in Virtual reality

Integrating virtual reality into the electronic commerce domain, several researchers [45], [46] suggest a completely virtual 3D e-commerce environment. Users are represented as 3D avatars that can walk in the environment and interact with products, sellers and other users. According to them, this kind of virtual reality shopping experiences creates convincing online virtual sensory affluence allowing the customers to grasp the value and product information more richly and engagingly.

With this approach, customers are allowed more control over the inspection of the products using the 3D models and visualizations. It has been found that giving control of information over to the consumer has major positive effects on their decision quality and motivation in electronic commerce, and generally results in higher confidence in judgment [47].

But the practicability of using it for a real business scenario is very limited, particularly due to the need of expensive virtual reality headsets and equipment. However virtual reality visualization might be suitable for some specific cases such as large-scale products like real estate and apartments [48] and luxury brands online trading [49].

Consumers of luxury products have various social, economic and psychological various intentions and motivations and seek to differentiate themselves from others and like to be treated uniquely. Providing a virtual reality based shopping experience facilitates these motivations by providing them a unique experience and is estimated to be an effective way to enhance the features of luxury products.

Altarteer et al [50] have researched on virtual reality real-time interactive product visualization and customization service where users can customize a product as they want while seeing and interacting with the resulting product in a realistic visualization. Their user studies have shown that users favor that system over traditional 2D image, animations, and video-based interfaces.

A less sophisticated version of virtual reality using mobile phones has emerged as an alternative to the expensive devices. There are some researches conducted on such mobile interactive virtual reality shopping environments [51] and user studies have shown that the usability and user experience is above average overall.

With all the potential of virtual reality, there are also some challenges that need to be addressed in order to ensure the success of VR in the e-commerce domain. Most significant would be the need for the expensive headsets with complementing powerful computers or smartphones. Most of the commercially available devices are too expensive for average consumers unless bought for another specific usage.

Effective usage of virtual reality in practice has limitations imposed by human sensory and motor physiology as well [52]. The human visual system is very sensitive to any irregularities between perceived visuals and other senses and

becomes prominent when motion is present in virtual reality. Any mismatches in motions and visuals can easily results in motion sickness in users within a short period and make the whole virtual experience uncomfortable.

Cyber sickness is a specific form of motion sickness syndrome that can arise as a result of exposure to VR and its symptoms include nausea, oculomotor discomfort, and disorientation [53]. Although specific causes of cyber sickness is not clearly identified yet, factors such as vection (the sensation of movement of the body in space produced purely by visual stimulation), lag, exposure time, field of view and individual causes are considered to be the reasons.

It is generally accepted that, visual senses suggesting that the body is moving and vestibular suggesting that the body is stable results in a sensory conflict which in turn results in cyber sickness. From a business perspective, cyber sickness occurring in a virtual reality based shopping experience raises liability issues for the retails or brands and poses a risk of making their customers uncomfortable.

Human years are capable of separating sound sources and approximating their directions, but in a virtual environment, it is challenging to provide a realistic auditory environment to facilitate this audio localization.

Virtual reality is yet to be available in mainstream usage and it will take some time for consumers to adopt and become more comfortable with virtual reality before retails and brands can invest in it as a profitable avenue. However not all the researchers agree on the use of virtual reality and some criticize it as just a tool that attracts users' attention rather than a solution for e-commerce.

2.1.6 Virtual Agents

Unlike in a physical service where there may be service agent or salesperson, human involvement in e-commerce is less from the service provider's side. Unlike physical stores that provide opportunities for social relationships such as interactions with salespersons and other shoppers, online stores are mostly anonymous automated

platforms, where there is hardly any social contact. But naturally, humans are more comfortable interacting and having conversations with another human. Offline shopping experiences involving positive social aspects have been shown to increase the time spent in the store, increase spending on products and increase impulsive purchases. But online stores face a challenge in introducing the social presence or the human warmth and sociability to online shopping experience [54].

Integrating a sense of human warmth and sociability can be done by providing actual interaction with a human or by stimulating the imagination of interacting with an imaginary person. In an e-commerce environment, actual human interactions may be provided through emails, sales support channels or forums. In contrast, there have been some efforts to introduce the human touch to e-commerce and other e-services by using virtual (imaginary) agents or assistants. The word agent is traditionally used to refer to a human who to provide guidance and similarly a virtual agent is an interaction metaphor to fulfill the sales agent's role with an anthropomorphic interface [55].

A number of researches can be found in the literature on experiments conducted with virtual agents with different characteristics and primarily these agents are of two types.

1. Completely imaginary agents with just interactions
2. Agents with a visual avatar form

Lucente has created a conversational natural language interface for e-commerce websites called "Expert" [56]. Users can have a natural conversation with the software agent as talking to a real sales person and it will answer users' questions and even suggest items to buy. They have found out that, users perceive "Experts" as having a personality. Another research [57] has shown that virtual agents are more effective when they are specialized in a particular category and enhance perceptions of information credibility, trust, and purchase intention.

Virtual agents with avatars, provides a combination of visual and or audio cues and allows the user to experience richer sensory stimuli and interactivity, and also increase the perceived social presence. Also, they realistically simulate or mimic the role of sales agents in online stores.

Qiu and Benbasat [55] have shown that consumers, who interacted with virtual agents with human avatars and voices, experienced a higher sense of social presence (or human warmth and sociability) than consumers who interacted with text-based virtual agents without an avatar.

Although not widely used practically, anthropomorphic virtual agents with multimodal interactions such as voice, gestures, and facial expressions are shown to have the potential of providing users with trust and positive experience [58].

Rigas, Dimitrios, and Nikolaos have conducted a research on using facial expressions and body gestures for agents with human-like avatars [59]. Majority of test subjects have reported a positive impression toward the agent and perceived by users as to simulate a real face-to-face conversation. These kinds of human-like avatars in e-Commerce provide different modalities to interact with the user and in combination with facial expressions and synchronized speech and gestures; it makes the characters more dynamic, more related and convincing for the users.

Trust and credibility is a critical factor for e-commerce system and some researchers have shown that virtual agents with avatars, particularly photo-realistic avatars with pre-recorded human voice enhanced credibility and trust in users considerably [60]. But according to another study [61], presence of talking avatars has negative effects on perceptions of information credibility and purchase intention among females, whereas among men the effects are positive. Their analysis suggests that females prefer textual content over auditory information spoken by the virtual agent about the products. Some experiments have also noted that users significantly prefer female avatars with gestures compared to male ones [26].

According to another research, virtual agents are particularly helpful for senior people with age-related disabilities [62]. They help overcoming social and psychological barriers elder people have for the adoption of online shopping. They have conducted research on 64 websites with virtual agents of different interaction styles (text, voice), realism (abstract, humanoid), gender and animations. They found out that, seniors least preferred agents with voice only interactions which may be due to the working memory problems they face with age. Also, they preferred agents with a male personality without a human-like avatar (abstract agent).

Although the interactivity of virtual agents is a key factor contributing to simulating social presence and it better simulates their roles as customer agents, some non-interactive virtual agents can also be found in use. Some websites utilize this kind of talking but non-interactive avatars that provide one-way communication in the forms of greetings, introduction or guides. Experiments [61] have shown that even non-interactive avatars like these improve the perceived social presence significantly. They can be a simpler alternative for the interactive virtual agents, which require more complex technologies, take time to implement and involve high costs. Third party services such as VideoPal [63] provide non-interactive talking avatars that can be embedded on any website. These services are convenient to integrate and also allow customizing the look and sound of the avatars as well.

2.1.7 Facial Expressions

In natural human communication, facial expressions play a huge role as a nonverbal communication medium. Facial expressions and other facial cues help deliver and understand the context of verbal communication even though we might not explicitly pay attention to them [64]. Often without the facial expressions, we misunderstand the context or meaning of electronically communicated messages. Trust is a key factor in interpersonal interactions. With facial expressions aligned with the verbal communication, it increases the perception of trust in the listener.

Inherently there is a lack of consumer trust toward e-services, especially when there are monetary transactions involved as in the case of e-commerce or e-banking. Thus HCI design for such services should consider building cognitive trust. Riegelsberger has conducted an extensive research on using facial cues to build trust and enrich user experience [65]. According to him, the advertising industry has been using photographs of faces to get an effective response and trust toward products. For example, an employee's photograph on an e-banking service will increase the feeling of trust in customers. His study on facial photographs on e-commerce sites found out that, it can build trust on some users as well as reduce usability for some others.

In another similar research, video clips and photographs of a presumed sales person was added to the home pages of e-commerce sites to simulate face-to-face interaction [66]. An extensive laboratory experiment has found out that the initial trust in consumers has indeed enhanced in those websites compared to websites without such images or videos. Also, the effect has been stronger when the depicted sales person was of the same geographical or racial origin as of the test subjects. It has been further proven by other researches that smiling expressions of model photos and intensity of smiles have an effect on consumers [67] and that those effects change depending on consumers cultural backgrounds [68].

Facial expression as a modality has been used with virtual agents as well. Lip-synchronization, eyes blinking, smiles can be incorporated into talking avatars to enhance the communication and to provide more realistic interactions.

Experimenting on this, Rigas and Gazepidis [59] has made a hypothesis that a human-like virtual salesperson with facial expressions, body gestures and lips movement synchronized to synthesized speech will have a positive effect on consumers. To test the hypothesis, they conducted an experiment with an e-commerce website where each product had a description, textual or spoken by a 3d avatar. Facial expressions and gestures were preset to the avatar for relevant parts of the descriptions. Majority of the test subjects rated positively for the products with

animated avatar description compare to the products with textual descriptions. 90% of the users agreed that such animated avatar can help them when shopping online.

Human facial expressions can also be used as an input modality in some cases. For example, lip-reading can be used to better recognize speech, eye movements, frowning can be used to detect attention and facial expressions that communicate feelings can be a measure of users' mental state.

It has been shown that in almost all cultures, six basic emotional facial expressions categories can be universally recognized. They are joy, sadness, anger, disgust, fear, and surprise [69]. Thus most of the facial expression recognition approaches focus on these six expressions. Still, automated recognition of facial expressions is a challenging task for computer systems.

There are two primary approaches to represent the face and facial features to and to analyze facial expressions [70].

- Geometric feature-based methods
- Appearance-based methods

The geometric facial feature-based methods represent the shape, texture and positional information of major facial features such as the mouth, eyes, nose, eyebrow, and chin, which varies from expression to expression. Thus this approach can cover the variations of appearance of facial expressions.

The appearance-based methods depend on image filter techniques such as Gabor wavelets to detect the facial feature for either the whole-face or specific regions in a face image. Noticeable points on the face, usually located on the corners of the eyes, corners of the eyebrows are generally detected as fiducial points in this approach. Using fiducial points to model the position of the major features, face geometry can be approximated in a local manner.

Emotion-oriented e-commerce is a new research field that utilizes facial expressions heavily. Emotions influence customers' shopping behavior and it is an important variable in purchase decision behavior. Nikov and Alexander [71] have presented a model for an intelligent emotion-oriented e-commerce system. In their model emotion recognition is based on facial recognition as well as on speech processing and other patterns. Such system, in theory, should be capable of identifying of non-purchasing emotions versus purchasing emotions and they can be used in the decision making to convert negative emotions to positive emotions which would lead to favorable purchase decisions.

2.2 Accessibility

New user interfaces and interaction modes are constantly researched and developed, but rarely do they take users with disabilities or elderly users into account. E-service developers are hardly aware of requirements of such users or do not see a business advantage in offering accessible services for them. But studies suggest that accessibility is indeed important for Business-to-Consumer (B2C) websites and it increases serviceability, reputation, and revenue of B2C [72].

Bocker et al. suggest that multimodality may contribute to providing accessibility in e-services [73]. They provide a set of design guidelines to evaluate how new interaction technologies will affect users with different abilities and how to overcome the accessibility limitations. In these guidelines, different characteristics of an interaction technology such as required user capabilities, user benefits, and cultural issues are considered.

Some researches have been conducted on developing age friendly e-commerce systems. One study [19] proposes to integrate crowd-improved speech recognition, multimodal search, and personalized speech feedback to facilitate the elderly's browsing of products in e-commerce environments.

Emmanouilidou and Kreps have developed a framework for implementing accessible services [74]. They have considered six different user groups with disabilities

including visually impaired, hearing impaired and motor impaired into the account. Their framework provides a detailed guideline on capability limitations of each user group and accessibility solution for them.

2.3 Context Awareness in E-Commerce

In the early days of computing, computers were large stationary objects, where it was set upped defined the context of use and it did not change. But with the beginning of the mobile computing that started to change. Users take their mobile devices with them and the operating environment and situations frequently change. Hence systems had to adapt to provide same functionality independent of the environment or improved functionality depending on the environment.

In his 1994 paper at the Workshop on Mobile Computing Systems and Applications, Bill Schilit introduced the term context-awareness and described it as systems' adaptability according to the location of use, the collection of nearby people, hosts, and accessible devices [75]. Early research on context aware computing focused more on location awareness but later researches identified context as any information that can be used to characterize a user and a situation [76]. Smartphone switching screen orientation or automatically adjusting screen brightness are two examples of how modern devices react to the context transparently to the user.

Although context can be defined by a number of factors, in practical use context is often hierarchically structured considering a limited set of relevant features or a feature space [77]. No feature space completely defines the situation and environment and it is adequate to have feature space that matches the system scope.

2.3.1 Personalized Services

Personalizing e-services based on the context of use is a highly researched area, particularly for e-commerce applications. Predicting user preferences and personalized suggestions, which are known as recommender systems, have been extensively deployed and become critical for e-commerce systems. It has been

studied that the recommender systems enhance e-commerce sales and increase customer loyalty.

Hong et al. propose an agent-based framework for providing the personalized e-services based on users preference history [78]. The framework has four layers where the first layer collects users' personal information and service access data with time, location etc. Collected context data is stored in the second layer and another layer extracts information from the stored context history. Application layer provides the personalized services based on the inferred information. Nevertheless, accurate recommendation and personalization is a challenging task due to complexities related to the capturing and use of contextual data and privacy concerns [79].

2.3.2 Location Aware Services

Going further, another research [80] has incorporated location awareness with the historical preference based e-commerce recommender system. According to them, with mobile devices being used to access e-commerce, an important piece of information, the user's location, is available. In addition to considering users' past preference, physical distances to the vendors are also considered to make a recommendation. There are other similar studies on location-aware mobile commerce [81] and location-aware augmented commerce [34]. In the later mentioned research they have experimented using user's location to render augmented reality content and provide navigation in an in-store e-commerce hybrid environment.

Location awareness which is a key driver and a critical success factor of the mobile advertising has been developed in drastically in the context of e-commerce marketing and advertising. The technology to pinpoint a mobile phone's location has a significant commercial value in the context of advertising. User attitude is yet again critical for location-based advertising. As users do not always want to announce their location and at the same time they need context sensitive information. Without careful control location-based advertising can become a form of spamming driving users away. B. Kölmel and S. Alexakis have introduced a scale named attitude

toward location-based advertising (ALBA) [82] to assist researchers in this field and check the openness of markets to location-based advertising.

2.3.3 Context Awareness with Multimodal Interactions

There have been some researches on using context awareness to combine multimodal interactions. SiAM-dp [83] is a platform for model-based development context aware multimodal applications. It supported creation of dialogue applications aware of and adaptive to the user, environmental context as well as communication modalities. Wolfram, et al. discuss a framework for context aware adaptation of mobile applications where user interface, content and interaction modalities adapt according to user context [84].

Context awareness has also been used in other e-services such as e-learning quite innovatively and some of these techniques are quite applicable in the context of e-commerce as well. In one research [85], they have designed a ubiquitous e-learning environment in a butterfly garden where plants are fitted with an RFID (Radio-frequency identification) tag. Students are given a mobile device with RFID reader. In this system, the computation is made to appear everywhere in the learning environment and they identify it as context-aware ubiquitous learning or u-learning.

Schmidt and Winterhalter [86] have presented a framework for delivering e-learning material in a context aware manner. They break down learning content into modular units and make them adaptable. Then they model the context of the learning environment and define the relationship between contexts and learning outcomes. Finally, the learner's current context information is used to find a matching learning objective.

2.3.4 Privacy Issues

Although context awareness can be used to provide personalized services/marketing which are appealing benefits to both consumers and businesses, it raises concerns regarding privacy [79]. To provide accurate personalized services such as

recommendations, systems have to collect a lot of personal data and consumers have to provide their private information expecting that they will get a better value or service and trusting the service provider will not disclose their personal information.

This belief of benefits, vary depending on whether consumers' preferences are discovered implicitly or explicitly. With the implicit discovery, there is usually a strong relationship between personalization and consumers' perceived privacy risks. On the other hand, explicit collection of consumer information does not raise consumers' privacy concerns and also increase the accuracy of recommendations, but requires more effort from the consumer. Thus, combined approaches are commonly used balancing user experience and accuracy of recommendations.

There are privacy issues also that are associated with location-based advertising. There are two forms of privacy issues, tracking and spamming. Tracking involves with people's unwillingness for their lifestyles being monitored by strangers. Spamming refers to sending unsolicited text messages to user's phone. Both these issues can be solved if the consumers are able to grant consent to advertisers [87].

3. METHODOLOGY

The purpose of this chapter is to show the research methodology of the thesis with the strategy of the research and the empirical techniques applied. First, we are going to analyze the importance of multimodality in e-commerce and related e-services and its impact on the overall quality of services. We will look at some of the preliminary studies conducted to study the effects of multimodality, and how based on the findings, interaction modes are selected to be supported by the framework.

3.1 Preliminary Study on Virtual Reality

With lack of hands-on experience with virtual reality, it was needed to conduct a preliminary study with virtual reality to its practicability and effects on users. But setting up a virtual reality experiment for an e-commerce scenario was practically difficult for various reasons including time constraints, resource limitations and need for 3d product models. Also, it was decided that this kind of a setup would not realistically simulate an actual online shopping scenario for users.

Thus, as an alternative, a preliminary study was conducted to study the effects of virtual reality (VR) and gesture interactions on users in another related e-service environment. E-learning scenario of a Massive Open Online Course (MOOC) was selected for this study due to the similarities it shares with e-commerce and ease of conducting a realistic experiment with available resources within a short period. In both e-commerce and MOOC based e-learning, services are accessed by a mass of general users and their interactions, user experience and engagement are critical factors for success. Thus it was reasonably assumed that, by studying the effects of virtual reality based interactions on user engagement and their user experience in a MOOC environment, we can get a reasonable understanding of its practicability and its effects on user engagement and user experience in an e-commerce environment.

The following sections describe this experiment and its results in details.

3.1.1 Background

Massive Open Online Courses (MOOCs) started to gain their popularity among resourceful universities and research institutes mainly due to the technological flexibility of hosting successful MOOC instances. This initiative created a positive influence on millions of students who could not access to learning in those leading universities; with the open nature of MOOC courses accessible online through browsers made these courses so popular among a range of prospective students.

MOOCs incorporate with a set of teaching and learner support tools and methods, which are meaningfully applicable to a range of different scenarios. It is important to note that MOOCs only offer limited types of learning options. Among these learning activities following are widely available in almost all MOOC platforms. Forums, Videos and streaming content, Quizzes, and Hypertext-based composed web pages.

While there are multiple success cases of MOOCs and their impact unquestionably significant towards making MOOCs a mainstream practice complementing the curricula, some of the critiques are of the view that MOOCs have yet to improve their learning offering to warrant as a formal educational methodology [88]. The main challenge with MOOCs is the poor completion rate and concerns have been arising on the real value behind MOOCs and the consequences of it.

Considering those limitations of MOOCs, the purpose of this experiment was to study how virtual reality based multimodal interactions can improve the user experience and how effective they can be for the learners.

3.1.2 Course Outline

For the experiment, a short course of Introduction to Bioinformatics was outlined and created to offer as a MOOC course. The unfamiliarity of the subject to the students, course time frame and the ease of visualizing biological concepts in 3D graphics, made this course an ideal candidate for the study.

Table 1: Structure of the Introduction to Bioinformatics MOOC course

Week	Topic	Unit	Component Type
1	Introduction to cell	Lecture	Video / VR
		Quiz 1.1	Problem (Checkbox)
		Quiz 1.2	Problem (Image Mapped Input)
2	Cell Division	Hand-out	HTML
		Visualization	Video / VR
		Quiz 2.1	Problem (Checkbox)
		Quiz 2.2	Problem (Checkbox)
3	Base pairs	Hand-out	HTML
		Interactive activity	HTML / VR
		Quiz 3	Problem (Dropdown)
4	Gene translation and transcription	Lecture	Video
		Quiz 4	Problem (Checkbox)

The course was designed for four weeks and the lesson content for each week consisted of a lecture and one or more pre-designed learning activities to cover the topics. The content of each lesson (see Table 1) was selected in a way to support the intended learning outcomes of the course module. The lesson for the 1st week was designed as an introduction to cell and cell structure, whereas the second lesson covered the concept of cell division and the different processes of mitosis and meiosis cell division. The third lesson explained the genetic material, DNA, their structure and how they are made up of matching pairs of nucleotides. This lesson was expected to be somewhat advanced for the students as they had not covered those things before and hence special care was taken to include an interactive activity to aid them. Lesson 4 explains how information contained in DNA is decoded by the process of translation and transcription and also how bioinformatics can be applied to solve various problems related to biological information.

3.1.3 The MOOC Environment

Although several MOOC platforms exist, Cousera and Edx are the most popular and widely used. None of the platforms currently supports VR content; therefore the selection for the experiment was solely based on the features available, the ease of deploying and the cost. After careful consideration, the free and open source MOOC platform, Open Edx [89] was selected to host the experimental setup.

The structure of the Introduction to Bioinformatics course was defined using sections, subsections and units and a unit contained multiple components of HTML content, problems and/or videos. The platform supported common problem types such as text input, dropdown and checkboxes, and advanced problem types such as image mapped input, which was used to create the quizzes for the course. The students were allowed to register to the MOOC and enroll in the courses and follow them. Through the LMS the instructors were able to access grades and interaction logs for each student and each component.

The course structure was defined to flavor and student-centered experience adding a more immersive experience, instead of forcing a sequential path as happen in traditional printed documents.

3.1.4 Virtual Reality Simulation Setup

To study the effectiveness of delivering lesson content as virtual reality experiences, virtual simulations were developed using Unity 3D game engine [90]. Unity 3D engine was specifically selected for the reason, that the created environment could later be exported to multiple platforms. Oculus Rift [40] was used as the rendering VR headset during the experimental setup, paired with Leap Motion hand and gesture tracking device [25] for hand input.

3.1.5 Experiment Design

Two courses were deployed on the Open Edx MOOC platform named Introduction to Bioinformatics as instructor-paced courses that runs for 4 weeks. Although the two

courses followed the same course structure (see Table 1), in one course, the activities designed for the first three weeks were delivered using virtual reality experiences whereas in the other course they were delivered through traditional learning methods. In both the courses, the lectures and activities in week 4 were delivered via traditional methods.

After enrolling in the course, the students can follow the course by going through the components for each section. In a MOOC the defined structure enforces only an implicit ordering for their guidance and the students are free to choose the order of topics and even navigate back to the completed components.

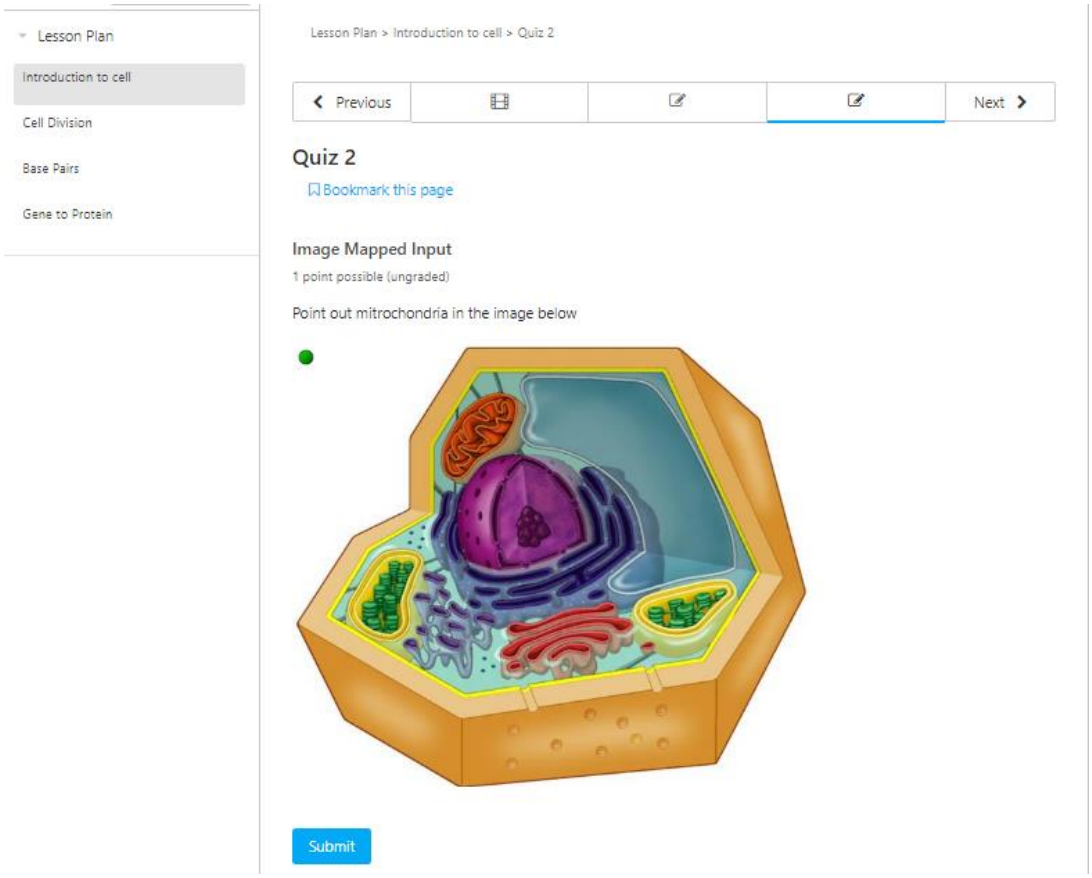


Figure 5: Student following a quiz attempt in the MOOC course

The three virtual reality experiences designed for the course had different objectives and different levels of interactivity. Lecture for week 1 was developed as a guided tour to introduce parts of a eukaryotic cell (see Figure 5). Throughout the tour parts

of the cell and their functionalities were introduced with voice-overs and annotated text. This experience was interactive and users could touch and move parts of the cell with their hands.

Visualization for week 2 targets meiosis cell division. The VR experience is a controlled 3D animation with a minimum level of interaction (See Figure 6). In contrast, in the interactive activity for week 3 users have to construct a DNA sequence with nucleotides (see Figure 7). The activity starts with a portion of a DNA sequence with base pairs with some nucleotides missing. Users were given a set of balls representing nucleotides and their task was to complete the base pairs by placing matching nucleotides at the correct sots. A basic guide on base pair matching was shown as an aid and only the correct nucleotide could be placed at each of the blank space.

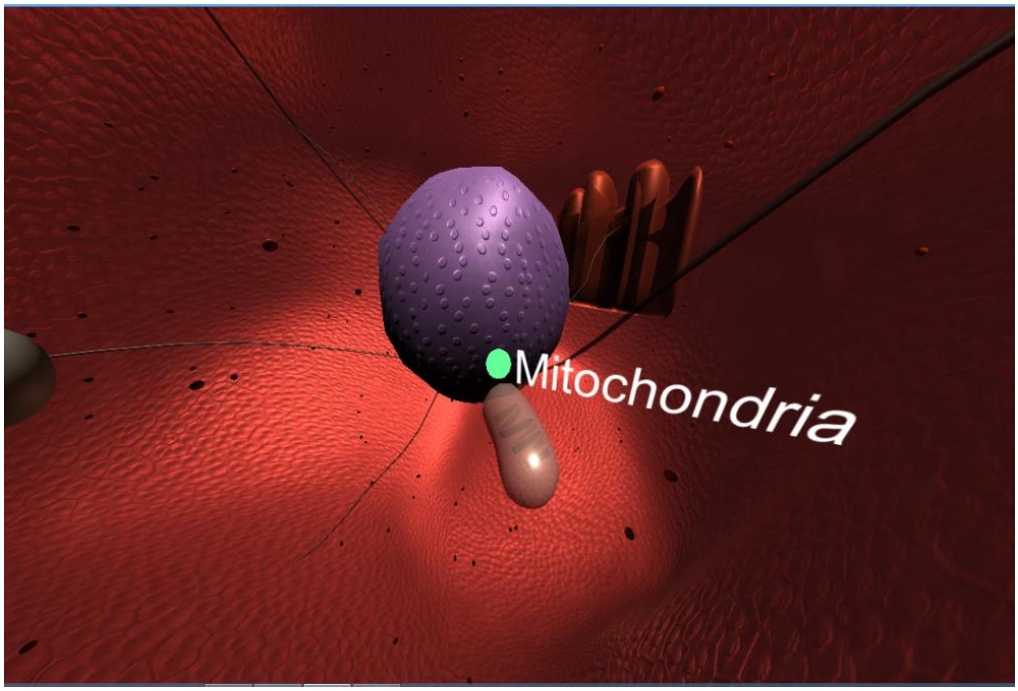


Figure 6: Interactive activity for week 1- Introduction to cell

In the other course, video content was available as lecture material and they were carefully designed to contain the same information as their VR experience counterparts. Lecture for week 1 was a rendered video of the eukaryotic cell tour VR experience with the same voice-overs. For week 2, lecture video consisted of a

rendered video of the cell division VR experience. For week 3, an online activity was provided in which students had to match nucleotides to complete base pairs of DNA sequence. Thus content wise both courses delivered the same information to the students.

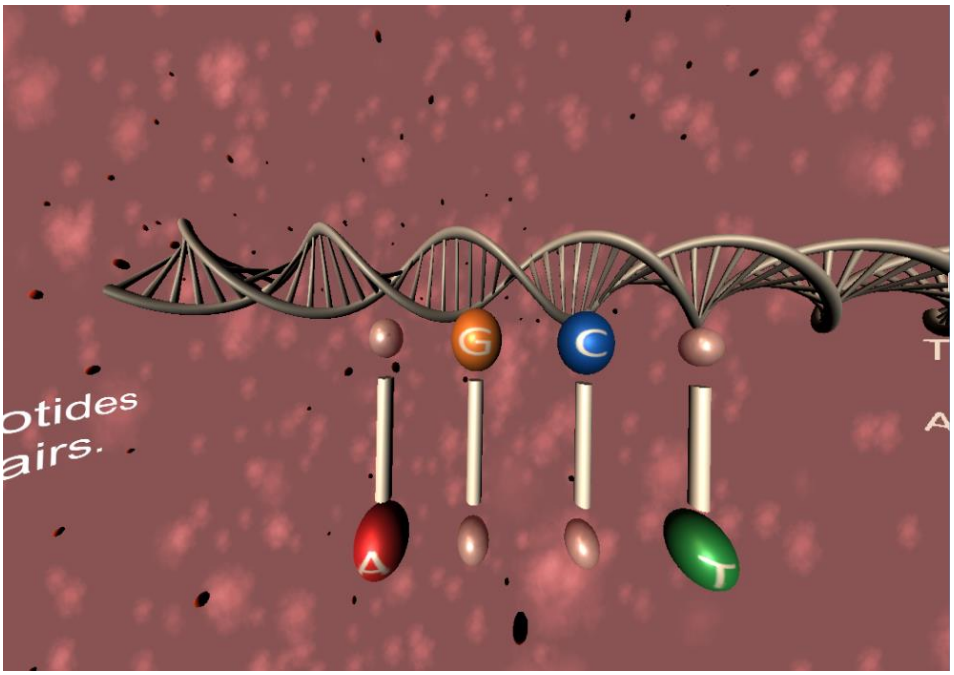


Figure 7: Interactive activity for week 3 – Base pairs

Two groups of students participated for the experiment with the system: 42 first-year undergraduates aged between 20 to 22 years and 42 final year undergraduates aged between 24 to 26 years who follow Bioinformatics course from Faculty of Engineering, University of Moratuwa, Sri Lanka. The goal of using these two groups was to distinguish between how the user experience and engagement varies among students who did and did not have prior knowledge on the subject taught depending on the method of content delivery.

Two sample groups were randomly formed in each of the previously mentioned student groups and they were asked to follow one of the two MOOC courses designed for the experiment. The test groups enrolled in the first course completed the relevant activities using the virtual reality setup (See Figure 8) discussed previously while the control groups enrolled in the second course and used the

traditional learning content. Each of the four small groups consisted of 21 students from each batch. The activity grades and time spent on each activity was collected as data from the MOOC platform to compare the effectiveness of learning and engagement.

Figure 8: A student completing the interactive activity in week 3 using VR

3.1.6 Results and Analysis

Figure 9 shows the average grade for the test group (VR) and control group (Video) from first year and final year batches. Since lesson 4 was delivered using the same content for control and test groups, comparing grades for Quiz 4, it can be established the performance of students in the control and test groups are similar within a batch.

By comparing lessons 1, 2 and 3, which were delivered using Virtual Reality and traditional content for test and control groups respectively, we were able to measure the impact of the learning material used on the student performance. Final year undergraduates who are following a course on Bioinformatics show better

performance when using traditional learning content. In contrast, the first year undergraduates have performed better when using virtual reality content.

The marks of the students in each activity were analyzed using analysis of variance (ANOVA) and some interesting results were found. The test group which had course content delivered in virtual reality showed significantly ($p<0.05$) better performance in quiz 1.1 with a mean value of 48.53% marks. In quiz 2.1 test group consisting of first-year students without prior knowledge in Bioinformatics, manage to score a significantly ($p<0.05$) higher mean score of 84.21% compared to the mean score of 50% of their counterparts who got accessed to the traditional learning content.

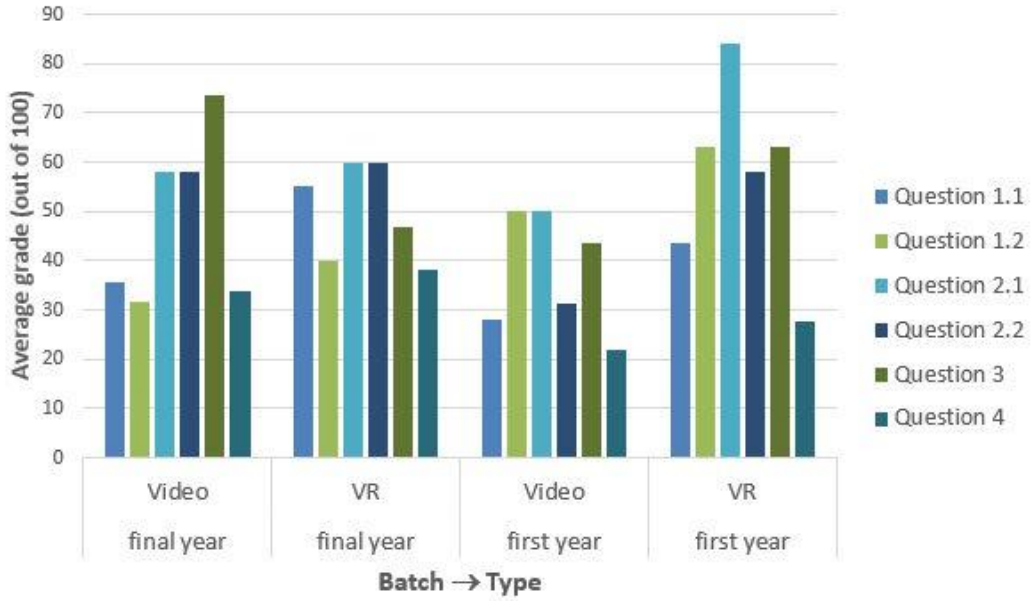


Figure 9: Average grade obtained for quizzes by the 4 student groups

The same control group had scored 62.36% in total for the first five quizzes which was significantly higher ($p<0.05$) than that of the test group consisting of first-year students. The last learning activity was not considered here as it was there as a benchmark and its lesson was same for both groups and in fact, the results confirmed that there is no significant difference of marks in quiz 4 between subject groups.

3.1.7 Student Feedback

In order to obtain a qualitative evaluation of the course content and the user experience, the students were asked to take part in an optional survey at the conclusion of the course. The first survey question was “Did you find the course content interesting?” (See Figure 10) shows the summary of student responses categorized by the study year if the students and the type of MOOC course they followed. The first year students express a clear preference for VR based content, while final year students display mixed responses.

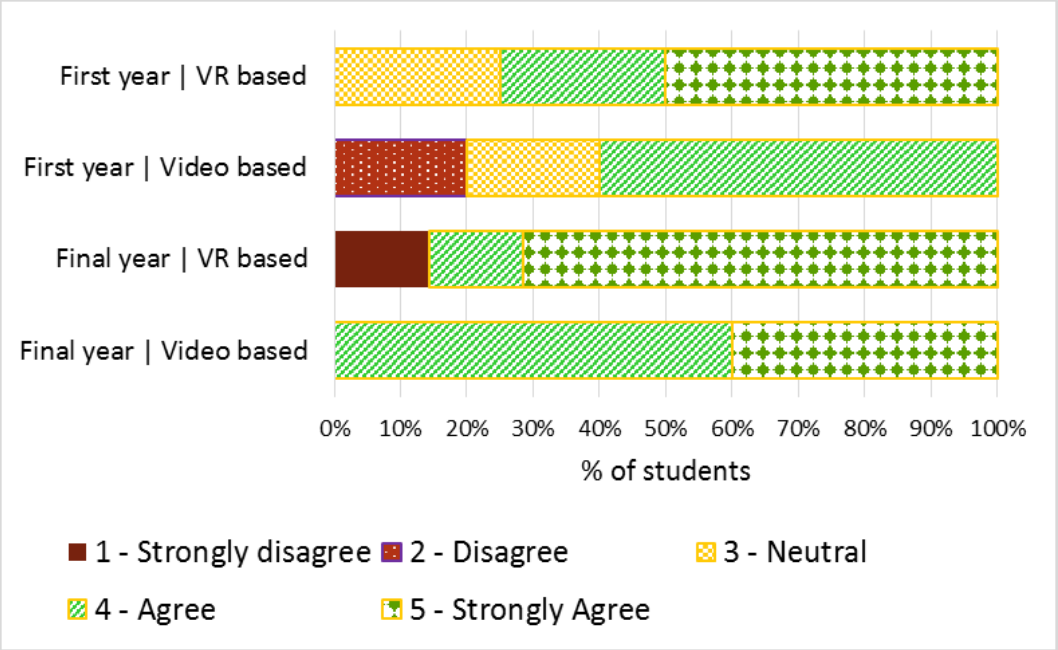


Figure 10: Rating on the interestingness of MOOC course content

Next, the students were asked to rate the likelihood of them following VR based and video based MOOC courses voluntarily in the future. Figure 11 shows the percentage responses of students in following video and VR based MOOC courses, categorized by the year of study. It can be observed that the final year students show a higher motivation to follow MOOC courses compared to first-year students, irrespective of the content delivery method. It is interesting to note that the tendency of first-year students to follow VR based courses and video-based courses are identical. Since this

contradicts with previous observations, we further analyzed the preferences of first-year students.

Figure 12 shows the percentage responses of students in following video and VR based MOOC courses, categorized by the year of study and the type of MOOC they followed for the experiment. It can be observed that the students who initially followed a VR based MOOC are more likely to try another MOOC course irrespective of the content type. Even the students who have followed a video-based MOOC initially shows a greater preference for VR based MOOCs.

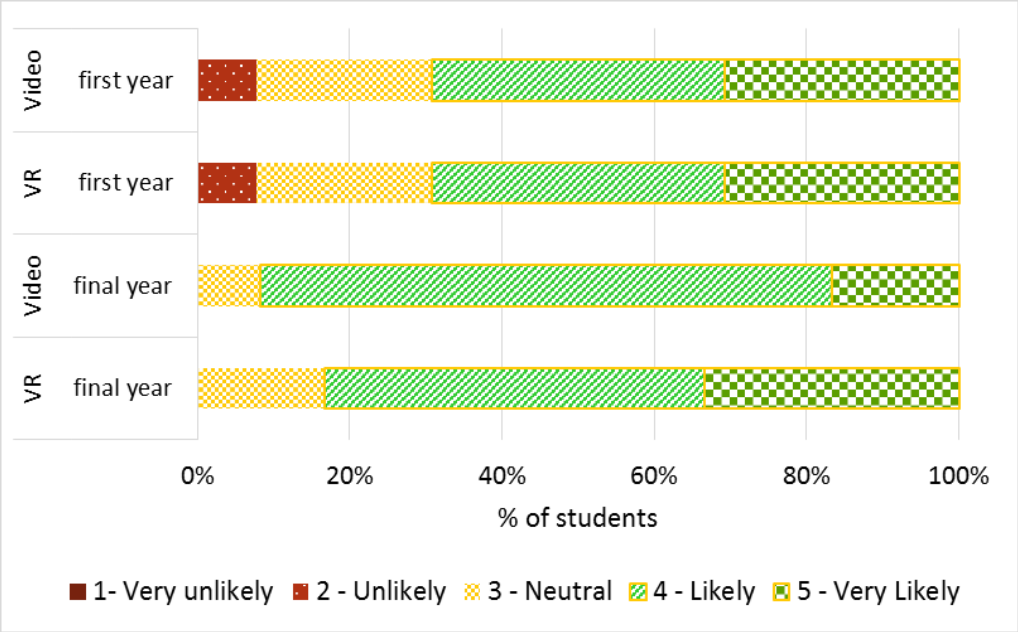


Figure 11: Likelihood of students following VR/ video based MOOCs

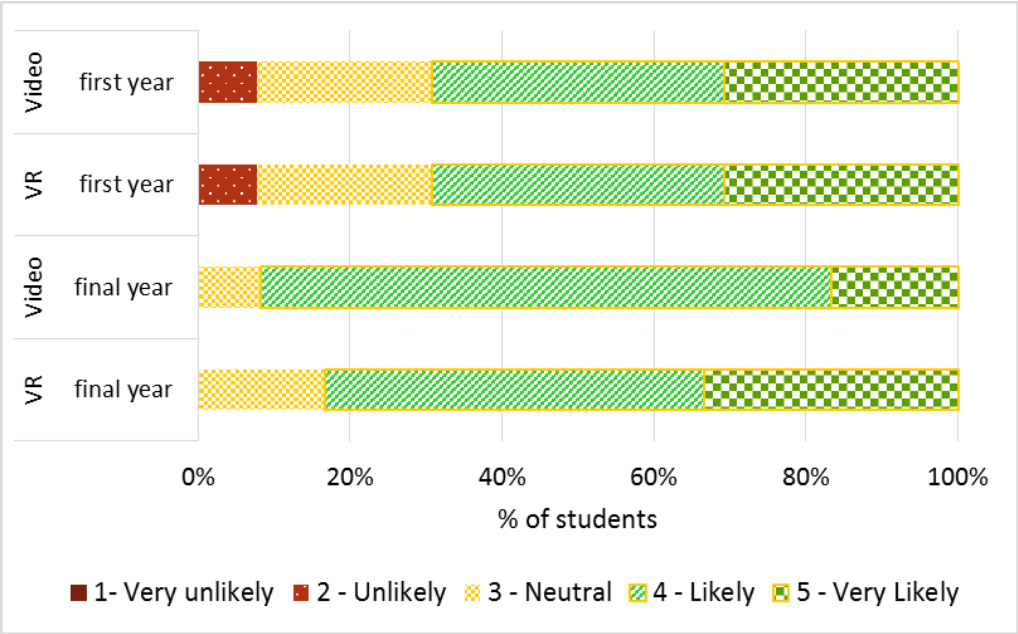


Figure 12: Likelihood of first-year students following VR/ video based MOOCs given the type of course they followed

3.1.8 Takeaway from the Experiment

It was hypothesized that the immersive experience from VR could be used to increase student engagement and user experience, than static forms of content such as video lectures. The student performance at each learning activity within the experiment MOOC and VR setup indicated VR simulations improve the students' learning experience within MOOCs. Another interesting observation was the student who had no prior knowledge on the subject matter gained higher levels learning success through virtual reality simulation supported MOOC compared to the students who had already followed a related course previously. Thus it can be assumed that virtual reality interactions are better suited for novice users rather than expert users.

From this experience it was evident that multimodal interactions particularly virtual reality and gesture-based interactions can improve user engagement with e-services and they provide a positive experience for them.

3.2 Identifying Suitable Modes of Interactions

This section describes about the selection of suitable interaction modes to be supported by the multimodal interaction framework, based on various factors. Although it is good as a concept, enabling multimodal interactions for actual e-commerce services is a challenging task. While they provide appealing capabilities to users, their value is often diluted by cultural, skill and context related hurdles [91].

3.2.1 Technical Challenges

Providing multimodal interactions for e-commerce platforms has some technical challenges particularly since most of the e-commerce systems are accessed through web browsers. Unlike a standalone application, a web application running on a browser has limited capabilities and resources.

Web Speech API [20] specification published in 2012 by W3C Community defined the JavaScript API to enable speech recognition and voice synthesis on web browsers. Still, only a few mainstream web browsers such as Chrome and Firefox support this specification, making it difficult to develop voice-enabled web applications for a wider user base.

There is no standard defined for the gesture recognition on browsers and whereas powerful computer vision libraries such as OpenCV [92] are available for desktop applications, no such advance libraries for web applications are available. To integrate gesture recognition for the prototype e-commerce platform for the experiment later mentioned in the paper, several open source JavaScript based gesture recognition libraries were tested. Most of them were found to be having limited capabilities and not practical to be used in real scenarios. Gesture recognition is a resource intensive task and JavaScripts running on a web browser only have access to limited resources, thus high performance cannot be expected from such a library.

For gesture recognition, the application requires access to the webcam of the user's computer. Although most of the laptops have inbuilt cameras, for desktop computers it is not the case. This considerably limits the scope of users who can access this modality. Also accessing the user's computer camera has some security concerns to cope with and most of the web browsers would allow that only for websites with secure origins.

Virtual reality particularly requires sophisticated and expensive hardware headsets such as Oculus Rift and high performing computers with high-end graphics processing units (GPUs). Thus only a narrow user base has the resources to experience virtual reality interactions provided in an e-service. Using a basic form of virtual reality tool such as Google Cardboard with smartphones is an alternative, but the quality of the experience will be lower and it still requires at least a mid-range smartphone with gyroscope and accelerometer sensors.

3.2.2 Socio-Cultural Challenges

It is a known fact that social and cultural backgrounds have an effect on how users perceive and interpret information [93]. The same multimodally presented information might not be understood similarly by two users of different cultural and geographic origins. In a system intended for a localized user group, this issue can be addressed by user evaluations. But for a system intended for a wider multinational or multi-cultural user group, it is practically difficult to evaluate beforehand.

In a research conducted to study how video clips and photographs of a presumed sales person in e-commerce websites simulate face-to-face interactions [65], the researchers have found that trust in consumers increase when the depicted sales person was of the same geographical or racial origin. It further shows how social factors affect the human perception.

Speech is the most natural, efficient as well as the diverse form of communication for humans. There are hundreds of languages in the world which have over million speakers each. Although voice recognition technologies have achieved remarkable

advancements in the last decade, their focus has been limited to widely used languages such as English, Chinese etc. and on native speakers of those languages. Most of the widely used languages such as English have many different variations of accent in different geographical areas as well as among cultural groups and non-native speakers of the language. These facts make providing speech-enabled interactions for an e-service intended for a multi-cultural and diverse language user base, extremely complicated.

3.2.3 Contextual Challenges

The context of a modern computing system is hardly a fixed one and this is especially true for e-services, where any users may access them from anywhere, any device and anytime. A modality that suits in a particular context may not be effective in another one. For example, speech interactions may work well in a low noise environment but may fail in a busy public location. Visual recognition based interactions are sensitive to lighting conditions and robustness is heavily reduced under low lights [22].

Age of the users also has effects on interaction modalities. Older users are less successful with motion and gesture-based interactions [94] due to their reduced motor functions and also speech recognition is less accurate for them with their reduced articulation skills [95]. Visual and hearing impairments are also common factors to consider with age and they impose major constraints on the interaction possibilities for those uses.

3.2.4 Cognitive Load

Although some interaction modalities such as speech mimic natural communication modes of humans and are intuitive for users, some other modalities have a cognitive load and a learning curve associated with them. 3D visualization is one such modality where the users have to perceive depth information presented on a 2D screen and interact with it.

When learning a new interaction mode the cognitive load required by the learner can vary and even exceed his abilities [96]. This mental demand depends on the person's age, gender, domain knowledge, and other factors and also a task that may require high mental effort in one user may not necessarily do so in another user [97]. For example, a user familiar with 3D designing software might find it more convenient to interact with a 3D visualization using a mouse rather than another modality.

In an e-commerce system where the general users are intended to access, it is impossible to assume the cognitive abilities of users at the development time, which varies drastically. Furthermore, it is important to identify and quantify the cognitive loads required by each modality before integration. But measuring the cognitive load is no straightforward task and requires complex physiological, performance-based, and self-assessment techniques [98]. Thus it is challenging to integrate interaction modalities that reduce the cognitive effort of users.

3.2.5 Selected Interaction Modes

Based on the literature review on section 2 and the results and observations from the preliminary study, the following interaction modes were selected to be supported by the proposed multimodal interaction framework for e-commerce. The challenges discussed in 3.2.1 were also considered for this selection.

- 3D visualization
- Gesture interactions
- Speech interactions

3D visualization was specifically selected for the interesting possibilities it allows in an e-commerce environment to visualize products in 3D. The literature review also suggested that 3d interactive product presentation generates a more positive attitude toward products and particularly the 360° views using the actual photos are more appealing to the consumers and give them a sense of presence. With modern browsers on desktop computers and mobile devices widely supporting 3D content

and graphics, 3D visualization was considered very much a feasible interaction mode on any platform.

Gesture interactions were selected since it is a natural way of interacting for humans and the novel experience it can provide for the users in combination with the 3D product visualization. Visual based gesture recognition could be used to achieve this since the majority of the laptops and smartphones have inbuilt cameras. Also, commercial products like Leap Motion were widely available in the consumer market for relatively low prices and seemed promising in providing accurate gesture recognition.

Speech is the most used natural interaction for humans and it was an obvious candidate for the supported interaction modes in the framework. Just as the cameras, majority of the laptops and all the smartphones have inbuilt microphones thus speech interactions are viable for a large user base. It had the potential to provide better usability, eliminating some of the typing uses have to perform to search for products in an e-commerce system.

3.3 Identifying Suitable Contextual Factors

Based on the literature review and observation made on existing e-commerce platforms the following contextual factors were selected to be supported by the proposed framework. Effect on the interaction modes and the interaction possibilities enabled by these factors were specially considered for this selection.

- Product
- User
- Device/platform of access
- Availability of hardware

The type of product has a direct impact on the suitable product presentation modes. Although it has been identified that, 3D visualization as a potential interaction mode, its applicability is limited to some types of products. For example, a relatively cheap item like a bar of soap or a food item may not worth presenting in 3D and seeing it in

3D will not add any value to the consumer. Thus product type and price are important contextual factors in determining the product presentation.

Most of the e-commerce systems rely on user information and history to provide customized services to individual users. The same concept could be used with the multimodal interaction framework, thus interaction modes can be prioritized and default interaction modes can be enabled depending on the users' past preference and interaction trends.

Not all the selected interaction modes were practical or applicable for all the platforms. It was determined that some interactions such as hand gestures would be difficult to perform on a mobile device. Thus the device or the platform of access was considered important as a contextual factor in determining the available interaction modes and also to configure them differently according to the platform. For example, a high-quality 3D product presentation on mobile devices would consume a considerable amount of mobile data and thus the quality has to be lowered.

Availability of hardware is a critical factor in deciding the availability of interaction modes as most of the selected modes require some sort of specific hardware interface such as a microphone for speech interactions, camera or Leap Motion device for gesture interactions etc. Unavailability of the required hardware features in the computer or mobile device can render these interaction modes unusable and thus it is taken as a contextual factor to enable, disable or customize the interactions modes.

3.4 Evaluation Scheme

After the implementation of the multimodal framework, it needed to be evaluated to measure as to how far it satisfies the intended objectives. For that an evaluation scheme was modeled.

The evaluation scheme consisted of two evaluation approaches: a quantitative analysis and a usability study. The qualitative analysis was intended to evaluate the

performance of the framework in terms of the positive influence it has on consumers and their purchasing behavior. Thus a simulated e-commerce scenario was planned and the e-commerce system for that has to be created using the implemented framework. Then the system was to be used by a test group of consumers, and it was planned to qualitatively measure the consumers' preferences changing some variables.

Similarly, for usability evaluation also a simulated e-commerce environment was to be created with the framework. A set of test users were to perform a set of predefined tasks on the system that covers all the interaction modalities supported by the framework, while being monitored. This approach was to evaluate the usability aspects of the framework.

4. IMPLEMENTATION

4.1 Design and Modeling of Framework

The architecture of the framework was carefully designed to cater to the basic requirements of an e-commerce system as well as the selected multimodal interaction capabilities. At a higher level, the architecture follows the typical Model-View-Controller (MVC) pattern.

Figure 13 shows the basic architecture design for the framework. Most of the e-commerce capabilities are in the models and controllers whereas multimodal interactions are mostly handled in the views.

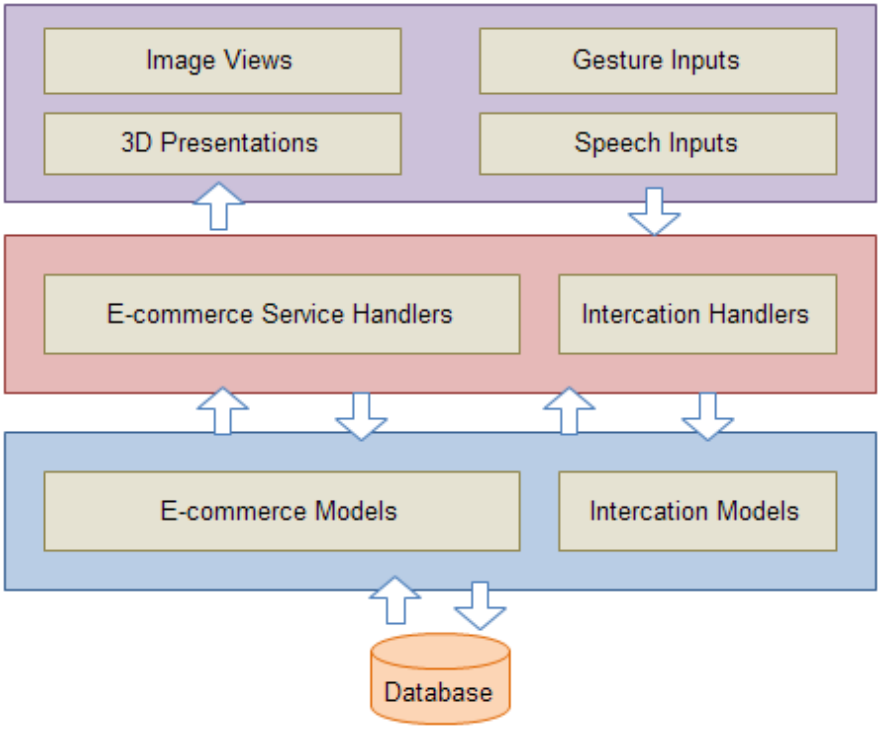


Figure 13: Architecture of the framework

Figure 14 shows the interaction architecture for the framework. Each input interaction was processed by the associated libraries or frameworks and the interaction manager integrates with the e-commerce context to provide the output

interactions. Implementation of each of these components is discussed in detail in the section 4.2.

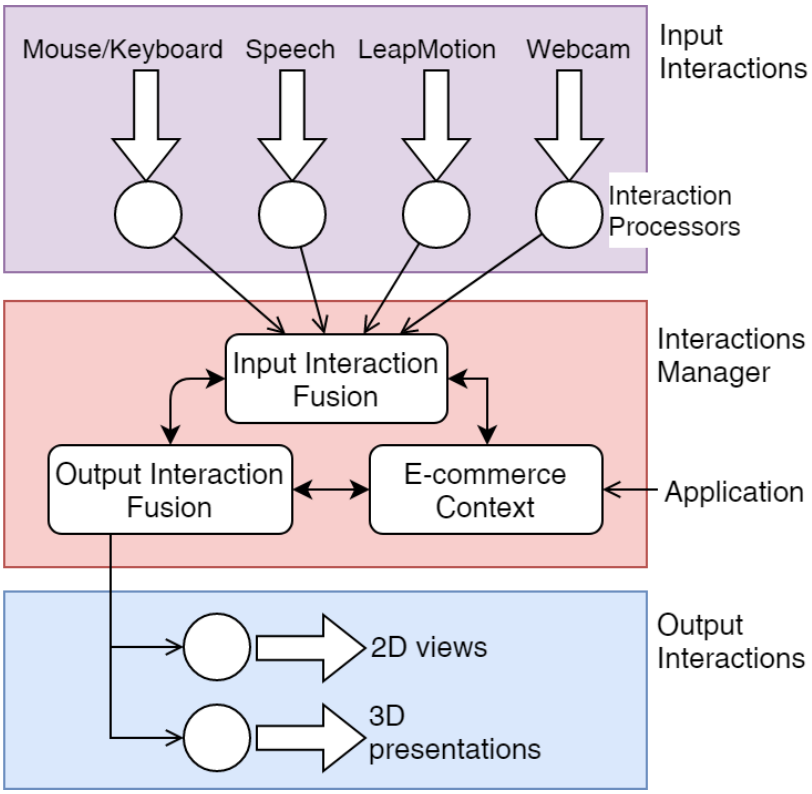


Figure 14: Interaction Architecture for the Framework

4.2 Development

The multimodal interaction framework was developed on top of the CodeIgniter [99] framework. CodeIgniter is a widely used open-source application development framework for PHP. It was specifically chosen for its model–view–controller (MVC) based design, small footprint, simplicity, and speed. CodeIgniter's latest stable version at the time 3.1 was used which supported PHP version 5.6 onwards.

E-commerce functionalities of the framework were implemented mainly on the server side adhering to the traditional MVC pattern while the multimodal interaction components were mostly implemented on the client side.

4.2.1 3D Product Visualization

The 3D product presentations were made up of a number of actual photos taken around the product in even angles and the resulting visualization could be smoothly rotated using the mouse giving the perception of a 3D object. Previous researches [28] have found out that the consumers have a more favorable attitude towards these pseudo-3D views made up of actual images, compared to 3D models, due to the realism and the interaction possibilities. The framework supported 12, 24 or 36 frames per 3D presentation taken around a single or two axes. With more frames more smoother the presentation to interact.

4.2.2 Webcam Based Hand Gesture Support

The framework supported interacting with the 3D visualization with mouse or hand gestures. For the hand gesture mode video stream from a web camera was used and its preview was displayed on the top corner of the page as a reference for the users. The gesture used for the interaction was moving the hand fist sideways and up-down in front of the camera. The 3D product presentation rotates according to movements of the hand. When recognized, hand fist was highlighted in the camera preview. Accessing and processing the webcam stream was implemented JavaScript using available open source libraries. As of yet only a few of the major web browsers allow accessing the webcam with JavaScript due to security and other concerns.

4.2.3 Leap Motion Based Hand Gesture Support

The framework also supported Leap Motion based gesture interactions to control the 3D visualization. Leap Motion is a hardware device capable of accurately tracking hand and finger using infrared light. The Leap Motion API returns the tracking data for a number of times per second as frames and these frames contain data about tracked entities, such as hands and fingers as well as approximated gestures (see Figure 15). In the case of a recognized hand, the following data could be acquired from the API.

- Palm Position And Velocity
- Direction And Normal Vectors
- Orthonormal basis
- Fingers
    - Tip position and velocity
    - Direction vector
    - Length and width

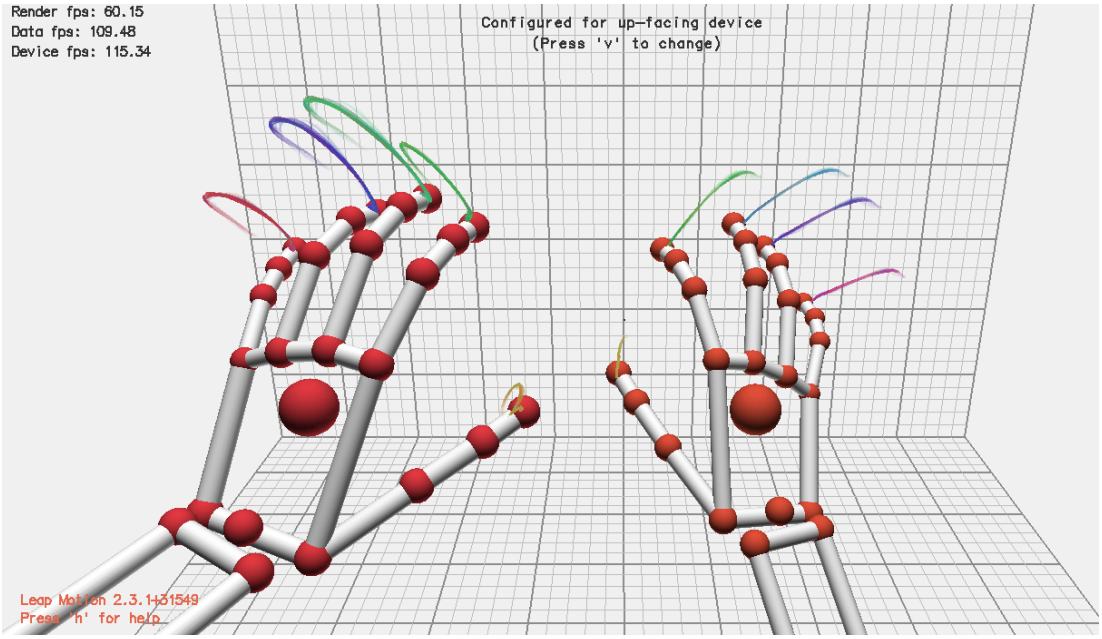


Figure 15: Hands and fingers as detected by Leap Motion API

With their JavaScript API, hand and finger positions, movements and speed data were monitored and using a simple algorithm rotation of the 3D presentation was calculated. This was implemented in a way that the gesture was only taken when the hand was fisted, so user could smoothly rotate the 3d product by moving a fist from right to left (or left to right) then opens the palm, and move back the hand to right and start moving again by fisting the hand. This was achieved using the grab strength property provided by the Leap Motion API which was an indicator for the palm is open or fisted.

The relative distance the fist has moved in horizontal and vertical directions were used to determine the angle of rotation for the 3D production view based on the

number of frames it had. It allowed controlling the 3D scene more smoothly and accurately by moving a hand sideways and up-down.

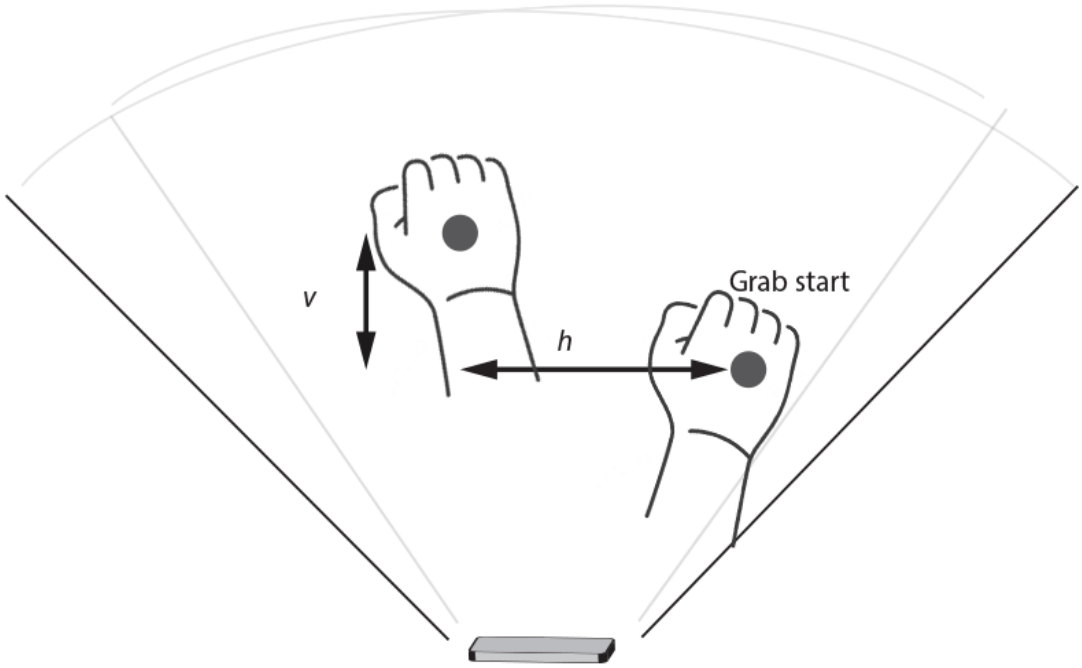


Figure 16: Hand movement used to rotate the 3D view

4.2.4 Voice Based Search

The search function in the framework was speech-enabled and users could press a button and speak what they want to search. Web Speech API [20] provided by the Chrome web browser was used to implement this and language mode was set to English (India) en-IN which was the closest for the English speaking style of the users. The native language of the users is not supported by the Web Speech API at the moment.

Although the Web Speech API specification has been published in 2012 by W3C Community, still only a few mainstream web browsers such as Chrome and Firefox support this specification, making this feature of the framework only available in those two browsers.

While speaking intermediate partial results of speech recognition were immediately displayed on the search input box so users get the feedback whether their words were recognized. Upon successful recognition, the framework automatically searched for products and displayed the results.

5. EXPERIMENTS

In this section, we present the experimental setup of two different experiments that were conducted to study different aspects of the developed multimodal framework for e-commerce. The aim of the first experiment was to study the effects of multimodal interactions on consumer preference whereas the second experiment focused on usability aspects of the framework.

5.1 Experiment Design

To evaluate the framework a case study was conducted by using the framework to develop an e-commerce system for a hypothetical scenario. The requirements for the hypothetical scenario can be summarized as below.

- Should support the general e-commerce functionalities
- Should provide a new experience for consumers
- Features to increase user engagement and purchase intentions

The requirements mostly aligned with the capabilities intended from the developed multimodal interactions framework. Thus the experiments were planned to evaluate how far these capabilities were achieved.

5.2 Effect on Consumer Preference

The goal of the first experiment was to study the effects of multimodal interactions in the developed framework, particularly 3D product presentations, on consumer experience, value perception, purchase decisions, and other factors. Furthermore evaluating how the methods of access (mobile, computer) or the quality of visualization affect this experience was also another aim of the experiment.

For the experiment e-commerce website developed with the framework was used. Six products were chosen including shoes, bags, and accessories in a way to balance the interested gender groups. With the same setup, two sub experiments were conducted as described below.

5.2.1 Experiment A

Each product was listed on the website without prices in three formats with the same textual description but with different product previews. In one format three images were shown as in a typical e-commerce website and in other two formats 360° turnable previews made up of 12 and 24 actual images were shown. The 24 images 360° preview was smoother to turn than its 12 image version. All the images were similar in size and textual descriptions were approximately similar in length. Brand names were removed from textual descriptions and images to eliminate the familiarity with the brand as a factor. Figure 17 shows an example product page used in the experiment.

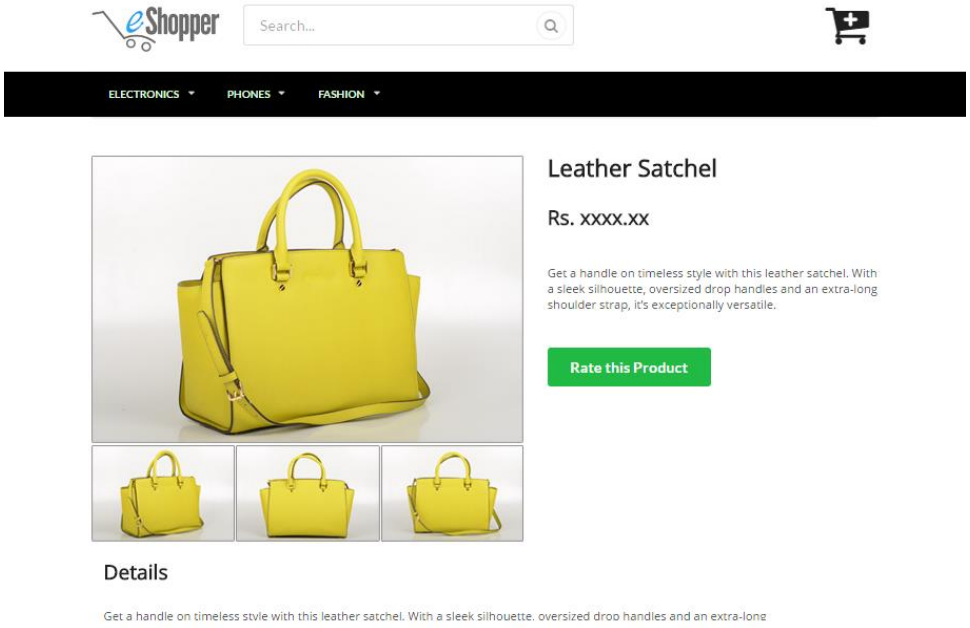
310 persons aged between 19 and 32 years participated in this experiment. 65% of the participants had some kind of previous online shopping experience. The participants were randomly assigned to three sample groups corresponding to the three product formats and were shown the six products in random order. The first group served as the control group and shown products with three image formats while the second and third groups were shown the product with 360° previews made up of 12 and 24 images respectively. 79 participants used smartphones to participate in the experiment while others used computers.

5.2.2 Experiment B

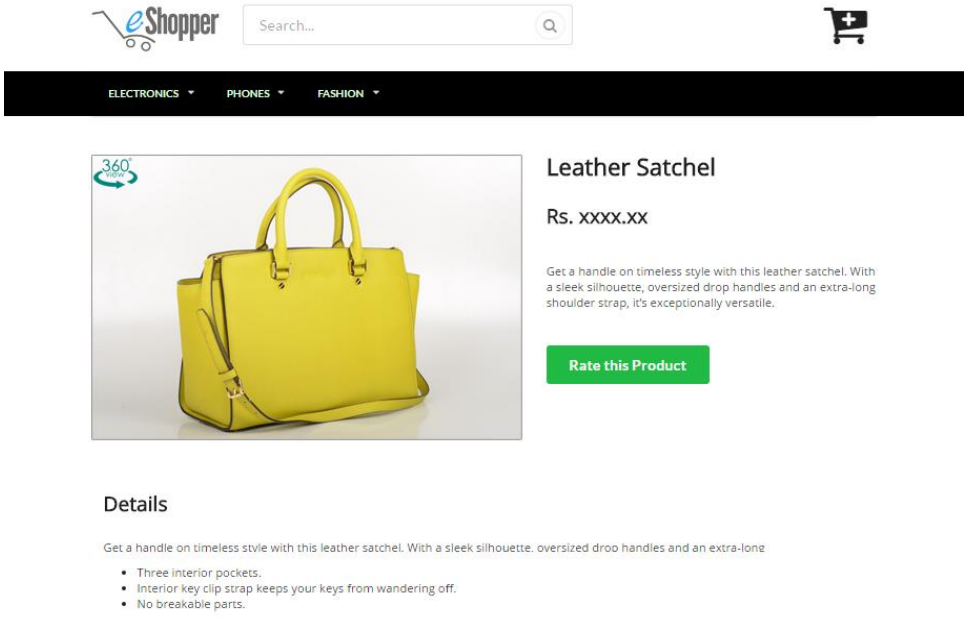
Similar to the first experiment products were shown in the created e-commerce website in two formats, one with three image previews and the other with 3D turnable preview consist of 36 actual photos. Unlike the 360° preview this 3D turnable preview allowed rotating the preview along both horizontal and vertical axes providing two degrees of freedom in control. Textual descriptions and other settings were as stated in the first experiment.

91 persons aged between 16 and 36 years participated in this experiment 67% of the participants had previous online shopping experiences. Similar to the first experiment participants were randomly assigned to two sample groups and shown products in the

corresponding format. 25 participants used smartphones to participate in the experiment while others used computers.



(a)



(b)

Figure 17: Screenshots of product pages of the e-commerce site created for the experiment. (a) Typical 2D images format (b) 3D interactive format

In both experiments, the time spent by each subject interacting with each product was measured and after viewing each product they were asked to report their opinion about the product by answering a set of questions. These questions covered attitude about the product presentation, perceived value of the product, likelihood of purchasing and the satisfaction level with the page loading time. To detect the perceived value, subjects were asked the maximum amount they would pay for each item. For the other three measures, their feedbacks on a 1-10 scale were asked.

5.3 Usability Evaluation

The goal of the second experiment was to study the usability aspects of the developed framework. For that, this second experiment was conducted as a formal user study in laboratory conditions. 15 test users (7 males and 8 females) aged between 21 and 28 were recruited for the evaluation. All of them were non-native English speakers and had previous experiences in using with e-commerce websites.

The subjects were given a set of tasks to perform with the system after a brief introduction to the system. In the first task, users were asked to use speech to search for products from a given set of keywords in English and they could attempt up to 10 times. The number of successful and failed attempts was recorded with the time taken for each attempt.

In the second task, users were asked to interact with the 3D product presentation using hand gestures as well as mouse or touchpad. Evaluation sessions were video recorded and monitored to find the difficulties faced by the users. Users were interviewed after the experiment and asked the following predefined set of questions to identify their opinions and difficulties they faced during the evaluations.

- What is your preferred method of interacting with a 3D product?
- What are the difficulties you faced interacting with speech, gestures or mouse?
- Are you comfortable using voice or gesture interactions when there are other people around?

6. RESULTS AND DISCUSSION

6.1 Effect on Consumer Preference

6.1.1 Experiment A

The results of the first sub experiment showed a significant ($p<0.05$) difference of attitudes toward the product presentation depending on the format products presented on a computer. It was found that participants have a favorable attitude towards products when they are displayed using a 360° view than the typical 2D image format. Interestingly, the favorable attitude was higher for 360° views made up of 12 images than 24 images.

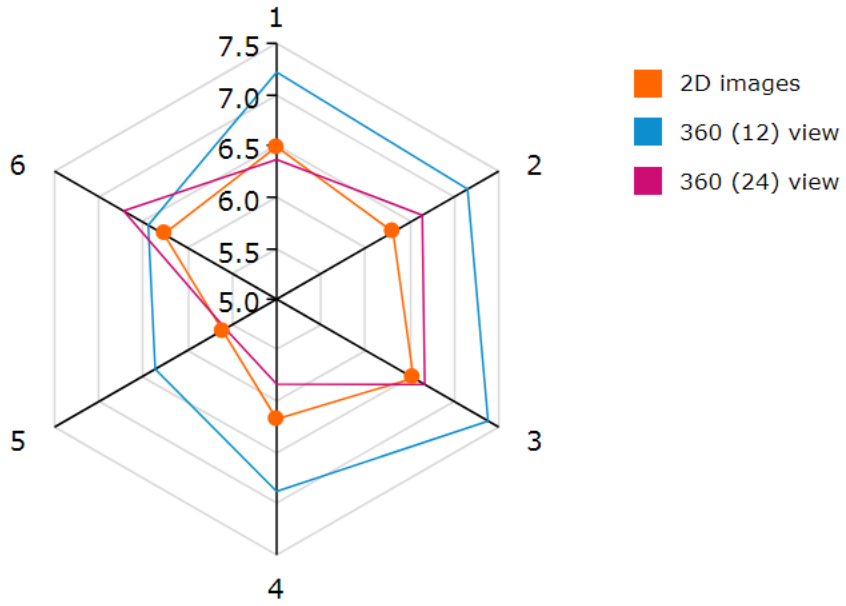


Figure 18: Comparison of mean consumer attitude of each product among 3 sample groups

In a similar pattern, the perceived value of all six products (see Figure 19) significantly ($p<0.05$) increased on computer for products displayed in 360° view with mean values of Rs.3338.55 for 12 image set and Rs. 2541.90 for 24 image set, compared to the mean value of Rs.2165.70 for the typical 2D image format. Most notably results showed that products presented in 360° view with 24 images significant increase the likelihood of purchasing compared to other two formats.

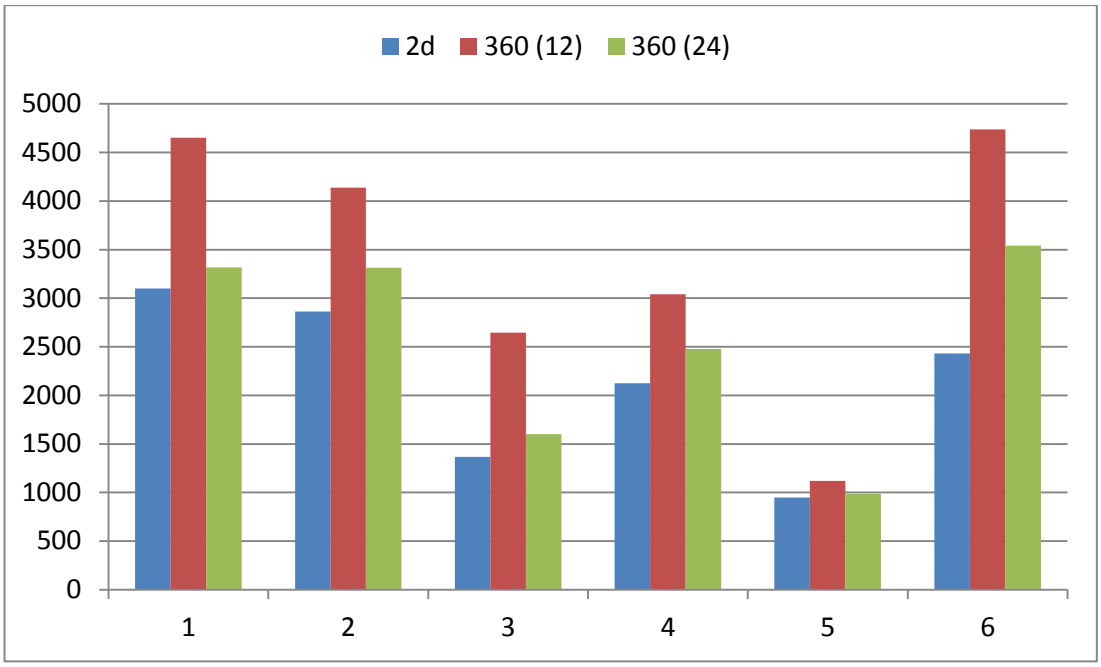


Figure 19: Mean perceived values of each product on computer in experiment 1

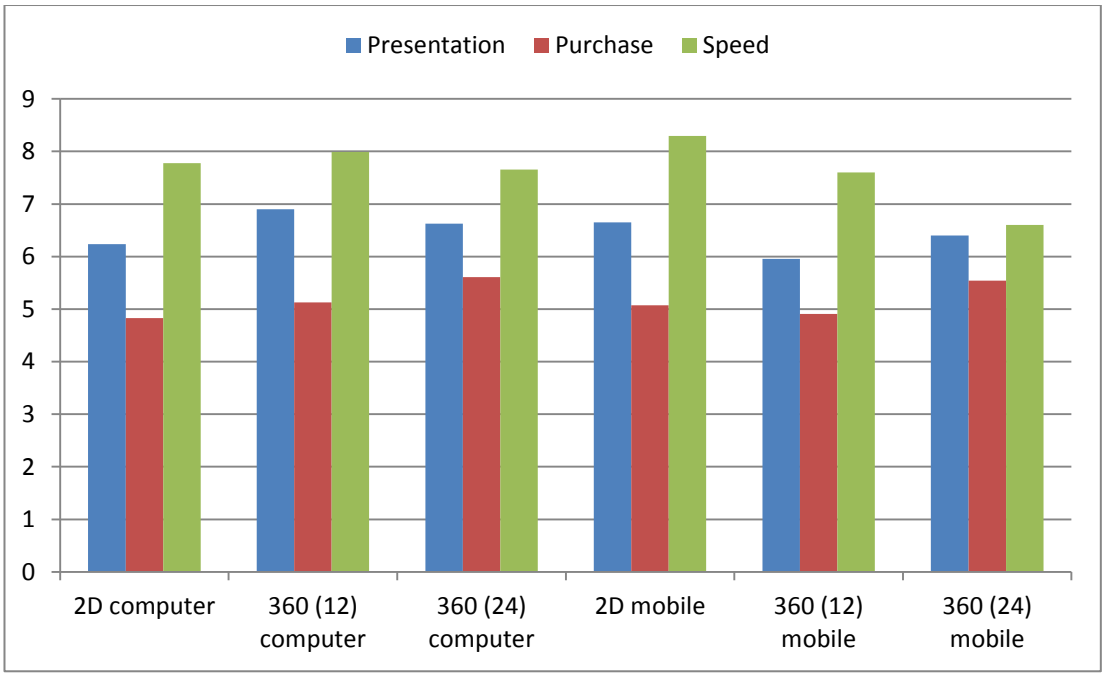


Figure 20: Comparison of consumer attitude toward product presentation, likelihood of purchase and loading speed between each test case

Although the actual page loading speed for products with 360° view is slower, the participants were more satisfied with the loading speed of the products with 12 images 360° view than products with normal presentation. This is probably due to the perceived satisfaction of seeing a 3D presentation without a noticeable loading delay. But 24 image 360° view showed a less satisfaction level in this case than the other two formats. No significant difference in the time spent engaging with products could be detected from the results.

Results from the participants using mobile phones showed a similar pattern in perceived value as in the previous case where a significant increase could be identified for the products presented in 360° view. No significant differences for attitude toward products or interaction time between the sample groups were found. But mobile users were less satisfied ($p < 0.05$) with page loading time of 24 images 360° views compared to the normal product presentations. 27% of the subjects were unsatisfied with the loading time.

6.1.2 Experiment B

From the results of the second sub experiment, it was found out that the 3D product presentation with 2 degrees of freedom in control did not improve the user experience or value of the product as expected on computers. In fact, the subjects who viewed the product in the 3D format have reported a significantly less likelihood of purchasing compared to the subjects who viewed products in normal format. Among the participants using mobile phones attitude toward 3D product presentation was negative. Still, the perceived value of products presented in 3D view showed a significant increase with a mean value of Rs.2672.08 compared to the mean value of Rs.1782.80 of products presented in normal format.

In this experiment, participants reported no significant difference of satisfaction level regarding the page loading time or time spent interacting with the product in both cases.

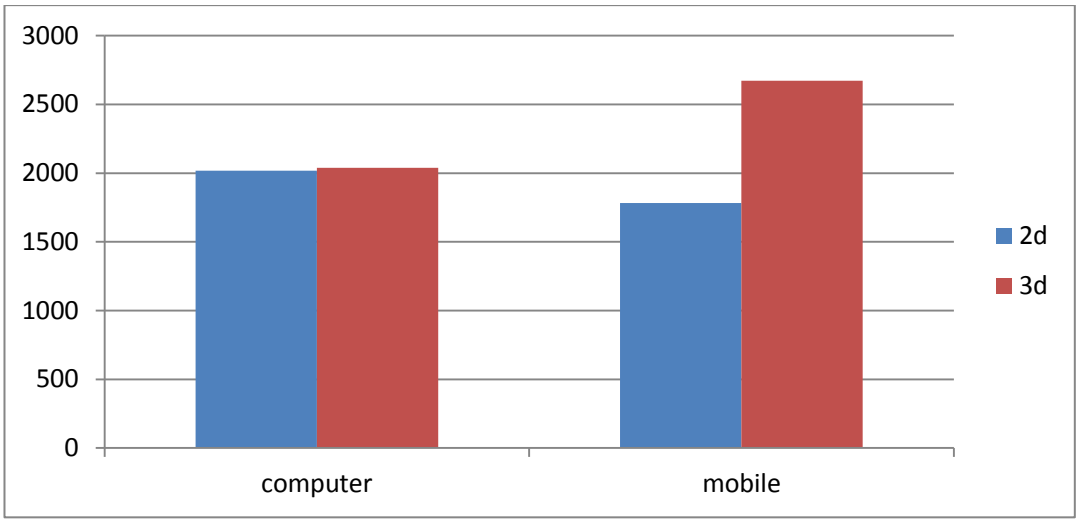


Figure 21: Comparison of mean perceived values between computer and mobile users in experiment B

These two experiments suggested that multimodal interactions in the framework, particularly 3D product presentations can have interesting positive and negative effects on consumer's shopping experience and purchase decisions. The results show that consumers have a favorable attitude toward 3D product presentation using 360° actual photos when they are viewed on a computer, agreeing with the results of previous studies [28]. This effect is also supported by the idea of perceptual curiosity [100], that is the consumers' desire for novel perceptual stimulation which results in the pleasure and emotional arousal which have a subsequent effect on purchase intentions.

Interestingly this study suggests that consumers perceive more value in a product presented in 360° view format and they are more likely to be purchased. What this implies is when a product is presented in such format with a price specified, consumers will feel like the price is cheaper and as a better deal, compared to the same product presented normally. Thus they are more likely to purchase the product. This is a key finding regarding the implemented multimodal e-commerce framework that can be used to improve the conversion rate of online shopping websites that can be developed with it.

Having smoother visualizations with more image frames shown to have mixed effects on consumers. Page loading time increases with the number of image frames and thus the decrease of satisfaction regarding the loading speed with smoother visualizations is understandable. Still, the positive effect of increased likelihood of purchase with smoother visualizations might worth the cost of slightly increased loading time.

Furthermore, a positive effect was expected from providing two degrees of freedom in controlling the 3D product presentation, but the results were implying quite the contrary on both computers and mobile phones. The reasons for this were unclear and it is possible that the difficulty of controlling the 3D scene with mouse or touch input is affecting the user experience. This was for some extent supported by the observations made in the usability evaluation as discussed in section 6.2.

However, the study indicates an increase in the perceived value of products with the degree of freedom provided on mobile phones. To make use of this positive effect, careful user experience design will be required to provide better interaction methods and a steep learning curve. The results also suggest that the loading time of 3D presentations is a concern for the mobile users. To reduce this overhead, image compression techniques can be used and the number of image frames in the visualization can be dynamically adjusted according to network speed. It is important to improve the consumer experience for mobile users as the percentage of mobile shoppers is rapidly increasing globally and about to surpass the percentage of desktop shoppers.

Although an increase in user engagement time with 3D presented products was expected, both the experiments show no significant difference. It is possible that as suggested in [101], allowing the product to be viewed in 3D results in fast reasoning and better understanding by the consumer thus reducing the engagement time. Further, observed experiments could be required to clarify that, but is out of the scope of this thesis.

6.2 Usability Evaluation

In the second experiment, all the users participated in the user study could successfully search using speech inputs and the mean success rate of a person per 10 attempts was 6.2 (s.d=2.14). The time taken for a single successful search varied from 4 seconds to 70 seconds with a mean value of 11.54 seconds (s.d= 12.99). Figure 22 shows the histogram for the distribution of time taken for successful search using speech.

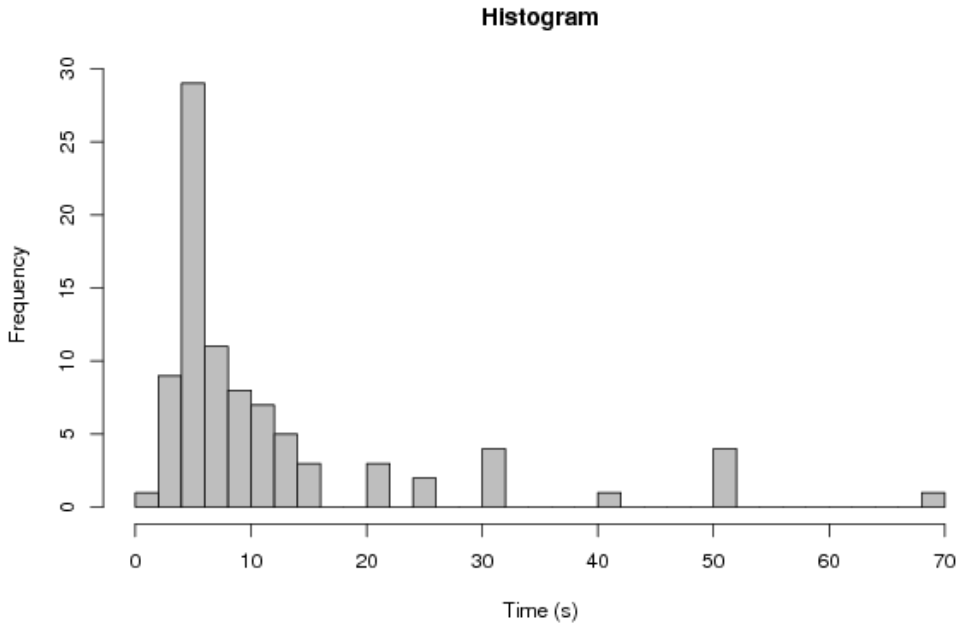


Figure 22: Distribution of time taken for a successful search using speech

In the second task, all the users were successful in controlling the 3d product view using mouse or laptop touchpads. But some difficulties could be identified when using the touchpad. 62.5% of the users (f = 10) reported that they were unable to use webcam based hand gestures to control the 3D view up to a satisfactory level. For the majority of users, their hands were not recognized by the system properly and the thus wasn't able to control the 3D view. On the contrary, all the users were successful in using Leap Motion based gesture interactions up to a satisfactory level.

In the post experiment feedbacks (see Table 2), only a minority of users (f=4) expressed that they would use speech or gesture interactions when other people are

around whereas 50.0% users (f = 8) said they would not and others (f = 4) were unsure. Majority of the users had reported hand fatigue as a difficulty they faced while gesture interacting (see Table 2), especially for webcam-based gestures.

Table 2: summary of qualitative feedbacks gathered from the interviews

Difficulties Faced	**Reasons for not Using in Public**
Speech	
Doesn't recognize my words Miss recognize my words Too slow	I might attract attention Might distract others
Gestures	
Doesn't recognize my hand Miss recognize my face as hand Difficult to control the view View rotates too fast My hand hurts Camera view is confusing Browser blocks webcam access	I might look ridiculous I might attract attention I might distract others

The user study revealed some interesting observation and challenges for the developed e-commerce framework and multimodal interactions in general.

Although all the users were able to use speech interactions successfully, some of them had some difficulties. A negative correlation could be detected between the success rate and the time taken per successful attempt. It is understandable as the less successful users had to try multiple times to get a correct search thus consuming more time. It was noted that users with clear and sharp voices were more successful

compared to the others. English was not the native language of any of the users and their accents were mixed and did not belong to any supported English modes. Some users were uncomfortable speaking to the computer while been monitored.

No specific difficulty could be identified nor reported by users while interacting with the 3D visualization using a mouse. When the other experiment on consumer preference showed some negative results in user experience when the visualization allowed two degrees of freedom in control, it was assumed that users find it difficult to rotate in two directions with the mouse. But we could not identify such difficulty in the user study. However, users who interacted using laptop touchpad, instead of the mouse, did face some issues. They could not rotate the product a full cycle in one touch motion and had to perform several touch-motion swipes to view the products' full 360 degrees. But with the mouse, the product could be rotated entire 360 degrees in a single motion.

Webcam based gesture interactions found out to be the most difficult for users and the majority of them could not even get their hands recognized by the system. Even though other users' hands were recognized for some extent, their success in controlling the 3D view with the hand was minimal. Backgrounds behind the users were noted to be largely affecting the performance of gesture recognition and when there was a monotonous and clear background behind, the gesture recognition was more accurate whereas with a cluttered background gesture recognition was extremely poor. Whereas powerful computer vision libraries such as OpenCV [92] are available for desktop applications no such advance libraries for web applications are available. Also, the fact that, a web application running on a browser has limited capabilities and resources makes it difficult to achieve a satisfactory level of visual gesture recognition.

Majority of the users reported hand fatigue after trying the gesture interactions for a small period of time. With the typical positioning of the webcam, users had to raise their fist to face level and keep unsupported for the duration of gesture interactions. Thus the fatigue in hands is understandable. This effect is well known as Gorilla Arm

Syndrome [102] and is a common phenomenon with vertical touch displays and hand gesture interactions. Holding arm in midair for gestures causes arm and shoulder fatigue in a short period and with prolonged durations, it results in physical discomfort and decreased performance [103]. There are recently developed matrices to quantitatively characterize this effect, such as Consumed Endurance [104].

When asked about using speech or gesture interactions while other people are around, the majority of the users have stated that they would not. Only 25% of the users have stated that they would use the multimodal interactions, while the others were unsure. They expressed that they will not be comfortable using these interactions in such environment and they might look awkward. Even during the user study, some users were not comfortable speaking and gesturing to the computer while being monitored.

This reveals an interesting social barrier for multimodal interactions. Shyness and social awkwardness are found all over the world but differ depending on the cultures and geographic regions [105]. Behavior in public can be viewed as acceptable, awkward or offensive depending on these factors. As evident in this study, users from some cultural backgrounds might find it uncomfortable to use some interaction modes in fear of unwanted attention and being negatively judged by others, in a social environment. This is a challenge for the developed multimodal e-commerce framework as the usage of its available multimodal interactions in a real scenario can be limited by this factor and thus its expected positive effects on consumer can get diluted.

7. CONCLUSION

E-commerce continues to grow as a major use of e-services with the widespread availability of the internet and computing devices. Although millions of users turn to online shopping on their ease from traditional in-store shopping, their shopping experience is limited by the lack of tangibility and unimodal interactions. As a solution, this research aimed to improve the e-commerce shopping experience by integrating multimodal interactions with context awareness.

Developing an e-commerce with such interactions and context awareness features can be a complex and tedious task. To simplify that task for some extent, this research designed and implemented a multimodal interactions enabled prototype framework for e-commerce, developed to improve the consumer experience. The framework supported three interaction modes: speech inputs, gestures inputs, and interactive 3D product presentations.

The developed framework was evaluated under two different criteria for its usability and the effect on consumers' preference in a realistic e-commerce scenario. In fact, it was found that the multimodal interactions allowed by this framework have some positive effects on the consumer and improve their consumer experience. Most importantly it has been shown that consumers perceive an increased product value, and they are more likely to purchase as well.

Resource limitations and lack of support for technology standards such as Web Speech API in web browsers had imposed major constraints on the capabilities of the framework. Still, the usability evaluation showed satisfactory results. Although webcam based visual gesture recognition in the framework found out to be less effective, Leap Motion was shown to be a good alternative for that with a good accuracy and lot of interaction possibilities.

Some usability issues could be noted in mobile devices. One challenge for mobile users to explore multimodal interaction views could be the lack of support they get for touch-based figure movements while holding the device at their hands.

Appropriate gestures that generate less fatigue can be used to let these users experience with the 360° view. With careful user experience design, those negative effects can be minimized to yield the positive effect of increased perceived value in consumers. Since mobile phones are becoming the dominant platform for online shopping, it is worthwhile to further research on that matter.

This research revealed some interesting social and cultural barriers to multimodal interactions that involve language and physiological factors as well. Some non-native English speakers shown to be having difficulties interacting with speech and enabling speech interactions for e-commerce systems with a diverse and multicultural user base can be extremely complicated. User acceptance of the developed multimodal interaction framework also found out to be affected by cultural shyness and social anxiety. As a result, users from some cultural backgrounds can be uncomfortable and hesitate from using some interaction modes while other people are around, in fear of being negatively judged.

However, this research has some limitations as well. Only a limited range of products have been used in the experiments due to lack of availability of 360 angle images. Due to that, this research did not cover how the observed effects of multimodal interactions enabled e-commerce systems would vary depending on the types of products.

7.1 Future Directions

From the findings of this research, we have identified several directions worth researching further.

We have highlighted some technical limitations and challenges that hinder the integration of multimodal interactions in the web context. These issues, from lack of support for specifications to lack of affordable devices, have to be addressed in order to multimodal enabled e-commerce or any other e-services to be effectively used by a mass of general users. Thus, we suggest that as a future research direction that needs attention.

This research revealed some interesting challenges to multimodal interactions that involve social and cultural factors. We suggest that further research focus should be on social psychological aspects of them and how these practical barriers can be overcome to bring multimodal interactions enabled electronic commerce platforms to a diverse user base.

The multimodal interactive e-commerce framework developed in this research paves way for further studies on how multimodal interactions affect consumers. It can be used as a platform for research, to study consumer dynamics by changing variables such as interaction modes, product categories etc.

8. References

[1] A. M. R. SSSD, “US Census Bureau Monthly & Annual Retail Trade.” [Online]. Available: https://www.census.gov/retail/index.html. [Accessed: 04-Dec-2017].

[2] C. Jeong, *Fundamental of development administration*. Selangor: Scholar Press, 2007.

[3] F. Corradini, A. Polzonetti, B. Re, and L. Tesei, “Quality of service in e-government underlines the role of information usability,” *Int. J. Inf. Qual.*, vol. 2, no. 2, pp. 133–151, 2008.

[4] T. Stivers and J. Sidnell, “Introduction: multimodal interaction,” *Semiotica*, vol. 2005, no. 156, pp. 1–20, 2005.

[5] J. J. Gibson, “The senses considered as perceptual systems.,” 1966.

[6] M. G. Millar and K. U. Millar, “The effects of direct and indirect experience on affective and cognitive responses and the attitude–behavior relation,” *J. Exp. Soc. Psychol.*, vol. 32, no. 6, pp. 561–579, 1996.

[7] J. Murray, “Composing multimodality,” *Lutkewitte Claire Multimodal Compos. Crit. Sourceb.*, 2013.

[8] W. J. Clara and C. S. C. Aileen, “An investigation on how designing video advertisements influences secondary school students’ perception of learner autonomy,” in *2013 IEEE 63rd Annual Conference International Council for Education Media (ICEM)*, 2013, pp. 1–16.

[9] A. Cangelosi, “Evolution of communication and language using signals, symbols, and words,” *IEEE Trans. Evol. Comput.*, vol. 5, no. 2, pp. 93–101, 2001.

[10] A. Jocuns, “Semiotics and classroom interaction: Mediated discourse, distributed cognition, and the multimodal semiotics of Maguru Panggul pedagogy in two Balinese Gamelan classrooms in the United States,” *Semiotica*, vol. 2007, no. 164, pp. 123–151, 2007.

[11] G. A. Hull and M. E. Nelson, “Locating the semiotic power of multimodality,” *Writ. Commun.*, vol. 22, no. 2, pp. 224–261, 2005.

[12] S. Oviatt, “Mutual Disambiguation of Recognition Errors in a Multimodel Architecture,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1999, pp. 576–583.

[13] J. Rouillard, “A multimodal E-commerce application coupling HTML and VoiceXML,” in *Eleventh International World Wide Web Conference, Waikiki Beach, Honolulu, Hawaii, USA*, 2002.

[14] Z. Trabelsi, S.-H. Cha, D. Desai, and C. Tappert, “A voice and ink XML multimodal architecture for mobile e-commerce systems,” in *Proceedings of the 2nd international workshop on Mobile commerce*, 2002, pp. 100–104.

[15] A. Acero *et al.*, “Live search for mobile:Web services by voice on the cellphone,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 5256–5259.

[16] E. Trentin and M. Gori, “A survey of hybrid ANN/HMM models for automatic speech recognition,” *Neurocomputing*, vol. 37, no. 1–4, pp. 91–126, 2001.

[17] D. Amodei *et al.*, “Deep speech 2: End-to-end speech recognition in english and mandarin,” in *International Conference on Machine Learning*, 2016, pp. 173–182.

[18] G. Hinton *et al.*, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[19] L. Meng *et al.*, “Towards age-friendly e-commerce through crowd-improved speech recognition, multimodal search, and personalized speech feedback,” in *Proceedings of the 2nd International Conference on Crowd Science and Engineering*, 2017, pp. 127–135.

[20] “Web Speech API Specification.” [Online]. Available: https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html. [Accessed: 24-Jun-2017].

[21] “Microsoft Speech Platform Overview (Microsoft Speech Platform).” [Online]. Available: https://msdn.microsoft.com/en-us/library/jj127858.aspx. [Accessed: 05-Dec-2017].

[22] S. S. Rautaray and A. Agrawal, “Vision based hand gesture recognition for human computer interaction: a survey,” *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[23] R. Nuwer, “Armband adds a twitch to gesture control,” *New Sci.*, vol. 217, no. 2906, p. 21, Mar. 2013.

[24] “Kinect Sensor.” [Online]. Available: https://msdn.microsoft.com/en-us/library/hh438998.aspx.

[25] “Leap Motion.” [Online]. Available: https://www.leapmotion.com.

[26] H. M. McBreen and M. A. Jack, “Evaluating humanoid synthetic agents in e-retail applications,” *IEEE Trans. Syst. Man Cybern.-Part Syst. Hum.*, vol. 31, no. 5, pp. 394–405, 2001.

[27] K.-S. Suh and Y. E. Lee, “The effects of virtual reality on consumer learning: an empirical investigation,” *Mis Q.*, pp. 673–697, 2005.

[28] F. Moritz, “Potentials of 3D-Web-Applications in E-Commerce - Study about the Impact of 3D-Product-Presentations,” in *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*, 2010, pp. 307–314.

[29] H. Li, T. Daugherty, and F. Biocca, “Impact of 3-D advertising on product knowledge, brand attitude, and purchase intention: The mediating role of presence,” *J. Advert.*, vol. 31, no. 3, pp. 43–57, 2002.

[30] M. Hassenzahl, A. Beu, and M. Burmester, “Engineering joy,” *Ieee Softw.*, vol. 18, no. 1, pp. 70–76, 2001.

[31] A. M. Fiore, J. Kim, and H.-H. Lee, “Effect of image interactivity technology on consumer responses toward the online retailer,” *J. Interact. Mark.*, vol. 19, no. 3, pp. 38–53, 2005.

[32] J. Kim, A. M. Fiore, and H.-H. Lee, “Influences of online store perception, shopping enjoyment, and shopping involvement on consumer patronage behavior towards an online retailer,” *J. Retail. Consum. Serv.*, vol. 14, no. 2, pp. 95–107, 2007.

[33] K. Khan and A. Nikov, “A mixed-reality-oriented eCommerce customer interface,” in *Proceedings of the 5th European conference on European computing conference*, 2011, pp. 152–157.

[34] E. M. P. Gunatunge, H. Y. M. Hewagama, S. G. S. Hewawalpita, I. Perera, and Y. B. N. Udara, “AugLAC #x2014; Generic framework for augmented location aware commerce,” in *7th International Conference on Information and Automation for Sustainability*, 2014, pp. 1–5.

[35] A. Mulloni, D. Wagner, I. Barakonyi, and D. Schmalstieg, “Indoor positioning and navigation with camera phones,” *IEEE Pervasive Comput.*, no. 2, pp. 22–31, 2009.

[36] B. Hettige, H. Hewamalage, C. Rajapaksha, N. Wajirasena, A. Pemasiri, and I. Perera, “Evaluation of feature-based object identification for augmented reality applications on mobile devices,” in *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*, 2015, pp. 170–175.

[37] J. Scholz and A. N. Smith, “Augmented reality: Designing immersive experiences that maximize consumer engagement,” *Bus. Horiz.*, vol. 59, no. 2, pp. 149–161, 2016.

[38] W. Zhu, C. B. Owen, H. Li, and J.-H. Lee, “Personalized in-store e-commerce with the promopad: an augmented reality shopping assistant,” *Electron. J. E-Commer. Tools Appl.*, vol. 1, no. 3, pp. 1–19, 2004.

[39] Q. Zhao, “A survey on virtual reality,” *Sci. China Ser. F Inf. Sci.*, vol. 52, no. 3, pp. 348–400, Mar. 2009.

[40] “Oculus Rift.” [Online]. Available: https://www.oculus.com/.

[41] “VIVE™ | Discover Virtual Reality Beyond Imagination.” [Online]. Available: https://www.vive.com/eu/product/. [Accessed: 27-Aug-2018].

[42] “Daydream.” [Online]. Available: https://vr.google.com/daydream/. [Accessed: 27-Aug-2018].

[43] G. Welch and E. Foxlin, “Motion tracking: no silver bullet, but a respectable arsenal,” *IEEE Comput. Graph. Appl.*, vol. 22, no. 6, pp. 24–38, Nov. 2002.

[44] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, “Perceptual Sensitivity to Head Tracking Latency in Virtual Environments with Varying Degrees of Scene Complexity,” in *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, New York, NY, USA, 2004, pp. 39–47.

[45] D. Contreras, M. Salamó, I. Rodríguez, A. Puig, and A. Yañez, “Supporting Users Experience in a 3D eCommerce Environment,” in *Proceedings of the XVI International Conference on Human Computer Interaction*, New York, NY, USA, 2015, pp. 45:1–45:4.

[46] K. C. Lee and N. Chung, “Empirical analysis of consumer reaction to the virtual reality shopping mall,” *Comput. Hum. Behav.*, vol. 24, no. 1, pp. 88–104, Jan. 2008.

[47] D. Ariely, “Controlling the information flow: Effects on consumers’ decision making and preferences,” *J. Consum. Res.*, vol. 27, no. 2, pp. 233–248, 2000.

[48] A. M. Paredes, “An examination of Mobile Augmented Reality Apps for the Commercial Real Estate Industry in Mexico City,” Master Thesis at London School of Business and Finance, 2014.

[49] S. Altarteer, C. Vassilis, D. Harrison, and W. Chan, “Product Customisation: Virtual Reality and New Opportunities for Luxury Brands Online Trading,” in *Proceedings of the 21st International Conference on Web3D Technology*, New York, NY, USA, 2016, pp. 173–174.

[50] S. Altarteer, V. Charissis, D. Harrison, and W. Chan, “Interactive Virtual Reality Shopping and the Impact in Luxury Brands,” in *Virtual, Augmented and Mixed Reality. Systems and Applications*, 2013, pp. 221–230.

[51] M. Speicher, S. Cucerca, and A. Krüger, “VRShop: A Mobile Interactive Virtual Reality Shopping Environment Combining the Benefits of On-and Offline Shopping,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 102, 2017.

[52] “Virtual Reality : Issues and Challenges.” [Online]. Available: http://web.tecnico.ulisboa.pt/ist188480/cmul/issues.html. [Accessed: 13-Sep-2018].

[53] N. G. Vinson, J.-F. Lapointe, A. Parush, and S. Roberts, “Cybersickness induced by desktop virtual reality,” in *Proceedings of Graphics Interface 2012*, 2012, pp. 69–75.

[54] K. Hassanein and M. Head, “Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping,” *Int. J. Hum.-Comput. Stud.*, vol. 65, no. 8, pp. 689–708, 2007.

[55] L. Qiu and I. Benbasat, “Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems,” *J. Manag. Inf. Syst.*, vol. 25, no. 4, pp. 145–182, 2009.

[56] M. Lucente, “Conversational interfaces for e-commerce applications,” *Commun. ACM*, vol. 43, no. 9, pp. 59–61, 2000.

[57] T. W. Liew and S.-M. Tan, “Exploring the effects of specialist versus generalist embodied virtual agents in a multi-product category online store,” *Telemat. Inform.*, Oct. 2017.

[58] J. Cassell and T. Bickmore, “External manifestations of trustworthiness in the interface,” *Commun. ACM*, vol. 43, no. 12, pp. 50–56, 2000.

[59] D. Rigas and N. Gazepidis, “A further investigation of facial expressions and body gestures as metaphors in e-commerce,” in *Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications*, 2007, pp. 148–153.

[60] A. ALVES and A. SOARES, “EVALUATING THE USE OF AVATARS IN E-COMMERCE.,” *Port. J. Mark. Port. Mark.*, no. 31, 2013.

[61] T. W. Liew, S.-M. Tan, and H. Ismail, “Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website,” *Hum.-Centric Comput. Inf. Sci.*, vol. 7, no. 1, p. 42, Dec. 2017.

[62] V. Chattaraman, W.-S. Kwon, J. E. Gilbert, and S. In Shim, “Virtual agents in e-commerce: representational characteristics for seniors,” *J. Res. Interact. Mark.*, vol. 5, no. 4, pp. 276–297, 2011.

[63] “VideoPal.” [Online]. Available: http://www.videopal.io/welcome/. [Accessed: 20-Sep-2018].

[64] C. Frith, “Role of facial expressions in social interactions,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 364, no. 1535, pp. 3453–3458, Dec. 2009.

[65] J. Riegelsberger, “The effect of facial cues on trust in e-commerce systems,” in *Proceedings of HCI*, 2002, vol. 2.

[66] K. Aldiri, D. Hobbs, and R. Qahwaji, “The human face of e-business: engendering consumer initial trust through the use of images of sales personnel on e-commerce web sites,” *Int. J. E-Bus. Res.*, vol. 4, no. 4, p. 58, 2008.

[67] Q. Wang, Z. Xu, X. Cui, L. Wang, and C. Ouyang, “Does a big Duchenne smile really matter on e-commerce websites? An eye-tracking study in China,” *Electron. Commer. Res.*, vol. 17, no. 4, pp. 609–626, Dec. 2017.

[68] Q. Wang, M. Wedel, and X. Liu, “How Facial Cues of Models Affect Attention to Websites in Asian and American Cultures,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2539253, Nov. 2014.

[69] T. Dalgleish and M. Power, *Handbook of cognition and emotion*. John Wiley & Sons, 2000.

[70] N. Sarode and S. Bhatia, “Facial expression recognition,” *Int. J. Comput. Sci. Eng.*, vol. 2, no. 5, pp. 1552–1557, 2010.

[71] S. Leon and A. Nikov, “Intelligent emotion-oriented eCommerce systems,” *J. Recent Adv. Artif. Intell. Knowl. Eng. Data Bases ISBN*, pp. 978–960, 2008.

[72] O. Sohaib and K. Kang, “The importance of web accessibility in business to-consumer (B2C) websites,” in *22nd Australasian Software Engineering Conference (ASWEC 2013)*, 2013, pp. 1–11.

[73] M. Böcker, H. Hüttenrauch, M. Pluke, A. Rodriguez-Ascaso, M. Schneider, and E. Zetterström, “Identifying enablers for future e-Services,” in *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, 2010, pp. 1–5.

[74] M. Emmanouilidou and D. Kreps, “A framework for accessible m-government implementation,” *Electron. Gov. Int. J.*, vol. 7, no. 3, pp. 252–269, Jan. 2010.

[75] B. Schilit, N. Adams, and R. Want, “Context-Aware Computing Applications,” in *1994 First Workshop on Mobile Computing Systems and Applications*, 1994, pp. 85–90.

[76] A. K. Dey, “Understanding and Using Context,” *Pers. Ubiquitous Comput*, vol. 5, no. 1, pp. 4–7, Jan. 2001.

[77] “Context-Aware Computing: The Encyclopedia of Human-Computer Interaction, 2nd Ed.,” *The Interaction Design Foundation*. [Online]. Available: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/context-aware-computing-context-awareness-context-aware-user-interfaces-and-implicit-interaction. [Accessed: 26-Jan-2017].

[78] J. Hong, E.-H. Suh, J. Kim, and S. Kim, “Context-aware system for proactive personalized service based on context history,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7448–7457, 2009.

[79] K. Verbert *et al.*, “Context-Aware Recommender Systems for Learning: A Survey and Future Challenges,” *IEEE Trans. Learn. Technol.*, vol. 5, no. 4, pp. 318–335, Oct. 2012.

[80] W.-S. Yang, H.-C. Cheng, and J.-B. Dia, “A location-aware recommender system for mobile shopping environments,” *Expert Syst. Appl.*, vol. 34, no. 1, pp. 437–445, 2008.

[81] X. Li, Z. Mi, Z. Zhang, and J. Wu, “A location-aware recommender system for Tourism mobile commerce,” in *The 2nd International Conference on Information Science and Engineering*, 2010, pp. 1709–1711.

[82] B. Kölmel and S. Alexakis, “Location based advertising,” *Mob. Bus.*, 2002.

[83] R. Neßelrath and M. Feld, “SiAM-dp: A Platform for the Model-Based Development of Context-Aware Multimodal Dialogue Applications,” in *2014 International Conference on Intelligent Environments*, 2014, pp. 162–169.

[84] W. Höpken, M. Fuchs, M. Zanker, and T. Beer, “Context-Based Adaptation of Mobile Applications in Tourism,” *Inf. Technol. Tour.*, vol. 12, no. 2, pp. 175–195, Apr. 2010.

[85] G.-Z. Liu and G.-J. Hwang, “A key step to understanding paradigm shifts in e-learning: towards context-aware ubiquitous learning,” *Br. J. Educ. Technol.*, vol. 41, no. 2, pp. E1–E9, 2010.

[86] A. Schmidt and C. Winterhalter, “User context aware delivery of e-learning material: Approach and architecture,” *J. Univers. Comput. Sci.*, vol. 10, no. 1, pp. 28–36, 2004.

[87] R. Unni and R. Harmon, “Perceived effectiveness of push vs. pull mobile location based advertising,” *J. Interact. Advert.*, vol. 7, no. 2, pp. 28–40, 2007.

[88] M. Bali, “MOOC pedagogy: gleaning good practice from existing MOOCs,” *J. Online Learn. Teach.*, vol. 10, no. 1, p. 44, 2014.

[89] “Open edX Portal | Open Source MOOC Platform |.” [Online]. Available: https://open.edx.org/. [Accessed: 25-Jul-2018].

[90] “Unity Game Engine,” *Unity*. [Online]. Available: https://unity3d.com. [Accessed: 25-Jul-2018].

[91] P. Carrara, D. Fogli, G. Fresta, and P. Mussio, “Toward overcoming culture, skill and situation hurdles in human-computer interaction,” *Univers. Access Inf. Soc.*, vol. 1, no. 4, pp. 288–304, 2002.

[92] "OpenCV library." [Online]. Available: http://opencv.org/. [Accessed: 24-Jun-2017].

[93] R. Nisbett, *The Geography of Thought: How Asians and Westerners Think Differently... and*. Simon and Schuster, 2010.

[94] A. B. Naumann, I. Wechsung, and J. Hurtienne, "Multimodal interaction: A suitable strategy for including older users?," *Interact. Comput.*, vol. 22, no. 6, pp. 465–474, 2010.

[95] D. Hawthorn, "Possible implications of aging for interface designers," *Interact. Comput.*, vol. 12, no. 5, pp. 507–528, 2000.

[96] S. Oviatt, R. Coulston, and R. Lunsford, "When do we interact multimodally?: cognitive load and multimodal communication patterns," in *Proceedings of the 6th international conference on Multimodal interfaces*, 2004, pp. 129–136.

[97] F. Chen *et al.*, "Multimodal behavior and interaction as indicators of cognitive load," *ACM Trans. Interact. Intell. Syst. TiiS*, vol. 2, no. 4, p. 22, 2012.

[98] E. Farmer and A. Brownson, "Review of workload measurement, analysis and interpretation methods," *Eur. Organ. Saf. Air Navig.*, vol. 33, pp. 1–33, 2003.

[99] "CodeIgniter." [Online]. Available: https://codeigniter.com.

[100] D.-M. Koo and S.-H. Ju, "The interactional effects of atmospherics and perceptual curiosity on emotions and online shopping intention," *Comput. Hum. Behav.*, vol. 26, no. 3, pp. 377–388, 2010.

[101] W. Schweibenz and F. Thissen, *Qualität im Web: Benutzerfreundliche Webseiten durch Usability-Evaluation*. Springer, 2002.

[102] T. Carmody, "Why 'gorilla arm syndrome'rules out multitouch notebook displays," *Wired Oct*, vol. 10, 2010.

[103] A. Jude, G. M. Poor, and D. Guinness, "Personal Space: User Defined Gesture Space for GUI Interaction," in *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2014, pp. 1615–1620.

[104] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "Consumed endurance: a metric to quantify arm fatigue of mid-air interactions," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 2014, pp. 1063–1072.

[105] K. H. Rubin and R. J. Coplan, *The Development of Shyness and Social Withdrawal*. Guilford Press, 2010.