Maximising the Transferability of Interaction Techniques for Immersive Technologies

Becky Spittle*

DMT Lab, School of Computing and Digital Technology, Birmingham City University, United Kingdom

ABSTRACT

A key factor encompassing immersive technologies is interaction, and the input methods employed to manipulate an interface. Nevertheless, barriers surrounding interaction and usability continue to hinder the development and potential of immersive technologies, with researchers having difficulties taking applications from laboratory environments to real use cases [5]. Therefore, the proposed PhD research aims to identify how natural input methods are best implemented in regards to commonplace tasks in immersive environments, and uncover how user approaches adapt depending on changes in environmental factors, the context of interaction and the hardware employed.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 INTRODUCTION

Interaction is an essential factor in immersive environments, 3D User Interfaces allowing tasks to be performed directly in a real or virtual 3D spatial context [1]. Interaction can be divided into implicit and explicit inputs, which refers to a combination of inherent motion and location awareness within the interactive space, as well as any intentional input provided to manipulate the scene [16]. Intentional inputs notably include selecting, translating, rotating and scaling virtual content within immersive environments [14].

When considering natural interaction, humans instinctively employ a range of senses and communication channels; exercising logic and considering context to create the most relevant interpretations, which permits real-time understanding [7]. A range of communication cues are used extensively in face-to-face collaboration, which includes aural cues; such as speech and para-linguistics, visual cues; such as gaze, gesture and facial expression, and environmental information; such as object manipulation, writing and drawing [3].

Simulating natural interaction, immersive technologies generally utilise complex systems that involve an amalgamation of user interaction, visual perception and other forms of multimodal communications [18]. Immersive applications can infer user input from the range of sensors employed, which are increasingly becoming built-in to consumer devices; posing opportunities to improve usability and accessibility, by creating more flexible, adaptive and inclusive technology ecosystems, which would benefit all users [6].

A balance of expressiveness and efficiency is required within all fields of human-computer interaction, but is arguably of higher significance in Augmented and Mixed Reality (AR/MR), as applications are broad and intertwined. These technologies often need to deliver interaction on the go, account for the danger of increased fatigue [20], and facilitate the need to seamlessly interact with both real and virtual content [4].

*e-mail: becky.spittle@bcu.ac.uk

3D interaction is therefore complex to employ, requiring novel configurations of interface components; such as devices, techniques and metaphors, which creates endless opportunities to design new, interactive methods [1]. Consequently, the range of existing and emerging immersive technologies currently adopt diverse interaction mappings and system architectures; the lack of standards resulting in a more staggered workflow for content producers and a less seamless and immersive experience for the user [5].

As a result, the long term goal of the proposed research is to formalise a framework for AR/VR interaction design, by strategically evaluating the interaction capabilities presented by immersive systems. This will be achieved by reviewing existing literature, and designing user studies that explore how situation, environment, use case and device affects interaction approaches. The proposed work will build on existing findings surrounding the appropriateness of different natural input methods, for a range of common tasks employed within immersive environments.

2 OBJECTIVES AND EXPECTED CONTRIBUTION

An opportunity to convey maximized transferability and interaction capabilities across immersive interfaces arises from multimodal input methods, as they provide a means to correlate proxies for natural interaction, across devices and use cases [10]. Although immersive technologies are beginning to become more practical for real use cases, most applications are employed sporadically and for ad-hoc tasks, whereas these technologies have the potential to provide continuous and multi-purpose user experiences [6].

Although input via additional devices can often be efficient and intuitive, due to familiarity with interaction through i.e. hardware-based control panels and remote controllers [20], the proposed doctoral research aims to understand how input methods can become standardised; and therefore focuses on input techniques that are achievable using the display device itself. As the most accessible AR/VR devices are mobile technologies, notably handheld (smartphone/tablet) and head-mounted displays (HMDs), input methods are based on what is intrinsic to these technologies using built-in sensors; such as microphones, cameras, gyroscopes and accelerometers. These vision and sensor based approaches interpret data for recognition and tracking, with primary inputs being speech, free-hand gesture, gaze/ head gesture and manual device manipulation; i.e. touchscreen input, or handheld motion gestures.

The primary aim of the research is to map the most appropriate input methods for different devices, tasks and use cases, and provide a set of guidelines for interaction designers and researchers. This involves addressing the following research questions:

1. How can multimodal, user-defined paradigms help to reduce system ambiguity and provide a more natural, synergistic and immersive experience? as well as:

2. To what extent can transferability of multimodal interaction capabilities be maximised across immersive experiences?

To generate answers to these research questions, and achieve the primary goal of providing researchers and developers with a

framework for design, the following objectives have also been defined:

1. To comprehensively review and evaluate how inherent interaction approaches are currently applied for immersive interfaces.

2. To design and conduct user studies that explore diverse inputs, tasks and use cases, to assess how multimodal communications could be mapped for everyday, meaningful interactions.

3. To interpret and review research findings and discuss how they relate and compare to current design and interaction practices.

4. To outline and develop a framework for interaction design in immersive environments, based on the advantages and disadvantages of multimodal input capabilities.

3 Related Work

The first task conducted as part of the PhD research was a literature review, which explored natural user interfaces and interaction approaches; as well as interaction for immersive applications. Following this, a more focused systematic review was performed that explored how different input approaches had been designed, applied and tested for explicit commonplace tasks in immersive environments.

Based on the two thorough, state of the art reviews conducted, key findings have been highlighted. The following subsections provide a breakdown of these findings, regarding input methods for different tasks, display devices and use cases. The literature reviews are informing the first user study, which is planned to be conducted in autumn 2021. Figure 1 provides a breakdown of the areas explored and introduces notable factors that have been considered to inform the research direction. Although technical areas such as recognition and tracking are essential considerations, the primary focus of the research surrounds how use case conditions affect interaction approaches and user experience.

3.1 Interaction

The main research topic explored is interaction in immersive environments, notably considering how input methods can be best utilised to deliver experiences that are adapted to the task being executed, the device being used and the context of interaction.

3.1.1 Inputs

The research primarily considers four types of input method, which are classified as; a) Free-hand: employing pre-defined gestures or unconstrained hand input with no wearable devices, b) Speech-based: using specific commands or natural language, c) Gaze/ Head gestures: interactions via eye gaze, head orientation/ rotation or head gestures, d) Manual interaction: where apparatus such as a touchscreen/touch-pad, or handheld device (phone/tablet) is employed.

It is understood in human interaction that hand and head gestures are not only instinctive, intuitively employed alongside speech as a fundamental component of communication, but that body movement is important to better understand and contextualise communication [15]. Although this knowledge has been applied to immersive environments, intuitive interaction technologies being highly considered in recent years, when considering the input and output modalities that enable interactions, some of the most natural forms of communication and information processing that humans possess; notably speech, language and gestures, are yet to be fully embraced [11, 12]. As well as this, there is currently a lack of guidelines and support to help researchers define interaction paradigms [5].

This has primarily resulted in input methods being implemented subjectively, with little consideration for alternative user approaches to interaction, which will likely generate interaction metaphors that are less natural and intuitive to employ [14, 17]. At present, interaction design introduces many complications and limitations, such as system learnability, and transferability (the provisions to interact on different platforms and in different contexts and settings), making it difficult to effectively implement novel systems and provide practical solutions for long term applications [5]. With this understanding, inputs are planned to be assessed to understand if there are any patterns in how users interact with a system in a range of environments and under diverse conditions, notably when different types of tasks are employed.

3.1.2 Tasks

Although existing, ad-hoc applications provide a range of interface utilities, input combinations and interaction methods, based on the work of Piumsomboon et al. [14], the types of explicit interactions have been classified into several categories; a) Pointing: defined as searching for interactive elements e.g. cursor or ray casting, b) Selection: initiating/confirming an interaction, c) Translation: moving or relocating an interactive element, d) Rotation: changing the orientation of an interactive element, e) Scaling: reducing or enlarging the size of an interactive element, f) Viewport control: zooming and panning within an environment via a specific function, g) Menu-Based: displaying a structured set of tabs, commands, and/or utilities for the user to interact with, and h) Abstract: interactions that could not be directly categorised as any other task, notably non spatial interactions such as editing (delete, redo, group and others [14]).

Regardless of the application type, often tasks in immersive environments are best performed using interactions that mimic how we perform activities in the real world. For example, we often use our hands to manipulate objects for everyday interactions, whereas because speech is naturally employed to communicate concepts, it is more difficult to apply for spatial interactions such as object translation, as it is difficult to precisely communicate intentions [19]. Instead, speech is notably beneficial for more abstract interactions, such as ”delete” and ”create” tasks; as it is more difficult to define gestures for non-direct, conceptual interactions [13].

Despite this, sometimes interaction preferences do not correlate with how we operate in the real world, and therefore designers are unable to simply apply our understanding of natural interaction to immersive technologies [5]. For example, scaling virtual objects is generally best achieved by employing interactions that are not permissible in a real-world context. Often paradigms that do not follow rules of human to human interaction are reused from familiar digital user interfaces, such as desktop based interaction with a mouse, or touchscreen gestures [11, 14]. As previously highlighted, speech can also be employed as a natural input method for straightforward, non-spatial property or function selections [19]. Although speech can not be used to directly enforce physical change in the real world, it is intuitive to employ as it is based on communicating definitive concepts; albeit to a digital system as opposed to another person.

However, although it may be more fun, intuitive, efficient etc. to employ a specific input type for a distinct task, other inherent factors must be considered; such as fatigue, novelty, social acceptance, and the hardware used [4, 5].

3.1.3 Display

Display device is a primary consideration, as the hardware used forces the user to interact via different methods. For instance, when using a mobile phone to interact with immersive applications, the user is generally required to employ at least one of their hands to hold the device (this most likely being both hands with tablet-based interaction) [8]. Therefore, although freehand gesture was found to be useful for object manipulation with HMDs [5], it is generally

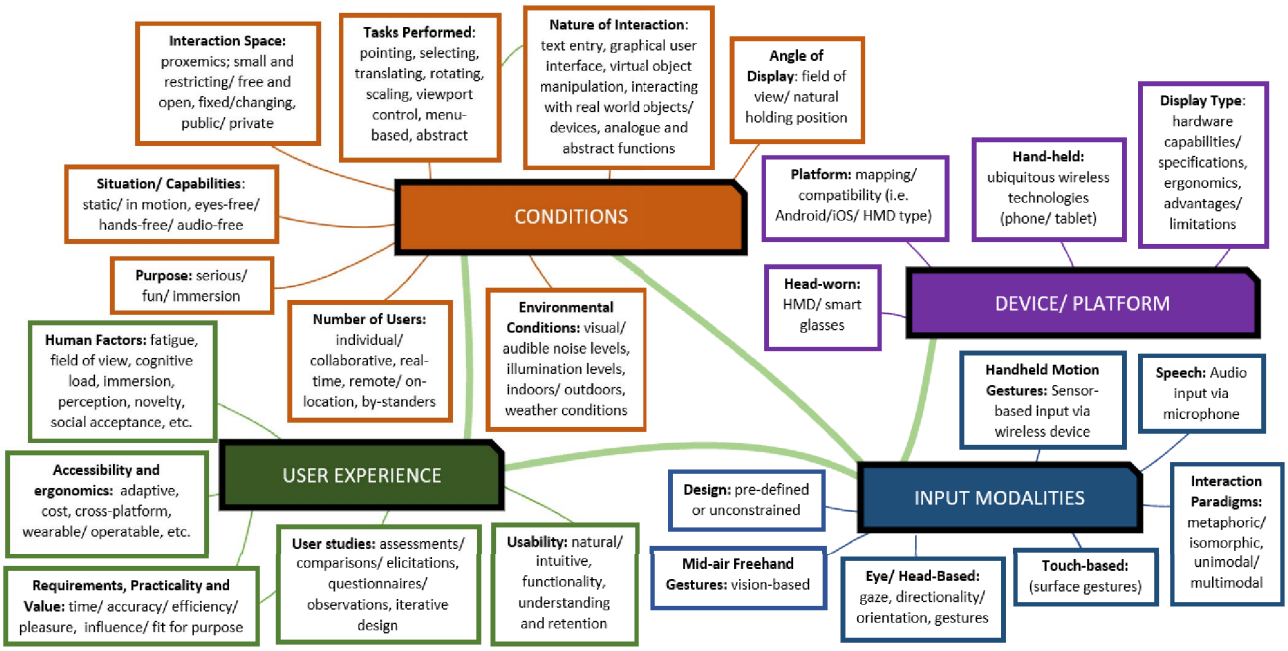

Figure 1: Mind map presenting the notable areas and factors considered when exploring interaction in immersive environments.

less appropriate for interaction with handheld display, especially without any hardware/software adaptions [8]. Head-based input was also found to be less accurate and natural when applied to hand-held devices, especially as distance increases [10]. Therefore, like hand-gesture based interaction, gaze and head input is predominantly more appropriate for head-worn displays.

There are several issues surrounding interaction with all mobile devices, notably in relation to ergonomics and technological constraints; such as tracking and recognition [9]. Field of View was also revealed to be a major factor affecting usability in both head-worn and handheld conditions, as well as depth perception and occlusion [8]. Consequently, by understanding what input methods perform well for the tasks classified on different platforms, we can begin to map input types and make it more straightforward to redesign applications for the different categories of hardware employed.

3.1.4 Use Case

Although the most appropriate input methods can be mapped to distinct tasks for individual devices, perhaps the most prevalent obstacle facing interaction with immersive technologies is the broad range of use cases and application types [4]. Use case refers to a number of factors surrounding the conditions in which an application is employed. This notably includes; a) the purpose of the application: if the interactions are serious or for fun, b) the situation of use: i.e whether the user has their eyes or hands preoccupied with another task, and c) the environment in which an application is used: considering how interaction methods are impacted by public/private settings, the size and nature of the interaction space (e.g. cluttered or open), as well as the level of lighting and audible/visual noise. By providing systems that are capable of sensing the current context of the user, immersive technologies can become pervasive; presenting opportunities to create continuous, adaptive, context aware experiences [6].

Often, user studies that assess interaction methods are conducted in a constrained lab setting. Rarely are 'in the wild' studies conducted that consider interaction approaches under realistic conditions. Although it is not always practical to conduct studies within a real use setting, the extent to which interaction approaches are pre-defined and restricted for the research will have an impact on the applicability of results [10]. An example of this is the lack of testing that assesses change in environmental conditions, as well as little consideration for how motion and portability of mobile devices is exploited by users; many researchers restricting users to interact from a single, static position.

As mobile devices provide opportunities to apply immersive technologies for a range of portable applications and scenarios [4], capturing information regarding natural user exploration and behaviours should be a key research consideration; however, this was found to not be the case. Few reviewed papers reported on movement trajectories or natural exploration and interaction approaches. Some studies attempt to simulate realistic conditions in a lab setting, however, the majority are highly controlled and restricted to a single condition.

Although measuring the same factors increases comparability, few studies considered more abstract measures; such as social acceptance and learnability, with few reporting on long-term studies or environmental factors such as noise or lighting conditions [9]. As a result, the findings generated are arguably less applicable to standard interaction applications and environments, which is potentially a factor hindering widespread implementation of immersive technologies for practical use cases [5].

4 Research Focus and Next Steps

Although natural input methods have been designed to provide a more engaging experience when compared to the otherwise sedentary nature of desktop environments [2], researchers are unable to simply apply understanding of the natural world to interaction with virtual content; user approaches differing depending on the task being performed, the nature of the interactive element and the situations they are interacting under [4, 5].

Consequently, to achieve the primary research goal of mapping inputs to different tasks; and provide a clear set of guidelines to interaction designers, more work is required to identify how the current performance of common input methods are effected by a

range of inherent factors, notably those surrounding the user, hardware employed and the context of interaction. By uncovering the impacts of less explored variables, such as those relating to the users activity and interaction environment, the framework can be applied and continually developed in line with the broad, expanding range of immersive technology applications [4].

A primary finding from the literature reviews conducted was the lack of consideration for how the situation, environment and use case may be effecting interaction approaches. Therefore, for the next stage of the research, a range of user studies are being planned to observe user preference for input approaches, when performing an assortment of tasks under different conditions. The first study is planned to consider different poses (sitting, standing, walking) and environments (public/private). Tasks will be delivered using a Wizard of Oz methodology, where scenarios will be presented to participants and they will be asked to demonstrate their natural approaches to input. Following user evaluations with a HMD device, the same studies will be repeated with a handheld device. Participants will be trained and introduced to the different types of input prior to conducting the studies, to provide inspiration for their interaction approaches [11].

Different factors surrounding the use case notably includes whether interacting indoors or outdoors and the level of light/noise, the crowdedness of an interaction space; in terms of the size of the environment and the density of surrounding people and objects (which can be measured subjectively or objectively), as well as factors surrounding the current state/ activity of the user. This final category relates to considerations such as the level and type of encumberment (i.e. number of hands occupied, and the types of objects being held), and the task scenario (whether interaction is associated with fun or serious applications).

The planned studies are expected to uncover how different interaction spaces, situations and environmental conditions affect preferred input methods for common tasks in immersive environments, the ways in which systems could measure and adapt to interaction preferences in different situations, how different input modalities can be employed to account for technical limitations in current systems; where factors such as recognition and tracking is affected by interaction conditions, as well as how multimodal interaction can be provided, without increasing factors such as cognitive load and learnability. By beginning to explore these areas through reviewing existing literature and conducting user studies, answers will start to be posed for the research questions defined in Section 2.

5 Conclusion

The overall goal of the proposed research is to map the most appropriate input methods for different devices, tasks and use cases, and provide a set of guidelines that outline these mappings to aid interaction designers. Based on findings from two literature reviews, the planned user studies have been outlined as part of this paper. These user studies are based on uncovering how multimodal, user-defined paradigms can help to reduce system ambiguity and provide a more natural, synergistic and immersive experience to users; by observing how interaction approaches adapt when immersive applications are applied in diverse situations and environments. By understanding the ways users adapt interaction approaches under diverse conditions, and with different display devices, we can also begin to reveal to what extent multimodal input capabilities can be used to maximise interaction transferability and user experience.

References

[1] J. Aliprantis, M. Konstantakis, R. Nikopoulou, P. Mylonas, and G. Caridakis. Natural interaction in augmented reality context. In *VIPERC@IRCDL*, 2019.

[2] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24:457–467, 01 2018.

[3] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI 2020, April 25–30, 2020, Honolulu, HI, USA, 04 2020.

[4] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan II. A systematic review of usability studies in augmented reality between 2005 and 2014. *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 49–50, 09 2016.

[5] M. Frutos-Pascual, C. Creed, and I. Williams. Head mounted display interaction evaluation: Manipulating virtual objects in augmented reality. *Human-Computer Interaction – INTERACT 2019*, 11749:287–308, 2019.

[6] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724, 2017.

[7] P. Jackson. Understanding understanding and ambiguity in natural language. *Procedia Computer Science*, 169:209–225, 2020.

[8] M. Kim and J. Y. Lee. Touch and hand gesture-based interactions for directly manipulating 3d virtual objects in mobile augmented reality. *Multimedia Tools and Applications*, 75:16529–16550, 02 2016.

[9] F. Manuri and G. Piumatti. A preliminary study of a hybrid user interface for augmented reality applications. *Proceedings of the 7th International Conference on Intelligent Technologies for Interactive Entertainment*, pages 37–41, 2015.

[10] S. Mayer, G. Laput, and C. Harrison. Enhancing mobile voice assistants with worldgaze. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 04 2020.

[11] M. R. Morris, A. Danielescu, S. Drucker, D. Fisher, B. Lee, c. schraefel, and J. O. Wobbrock. Reducing legacy bias in gesture elicitation studies. *interactions*, 21:40–45, 05 2014.

[12] C. Munteanu, K. Nakamura, K. Nakadai, P. Irani, S. Oviatt, M. Aylett, G. Penn, S. Pan, N. Sharma, F. Rudzicz, and R. Gomez. Designing speech and multimodal interactions for mobile, wearable, and pervasive applications. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, Association for Computing Machinery, New York, NY, USA:3612–3619, 2016.

[13] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 09 2014.

[14] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. *Human-Computer Interaction – INTERACT 2013*, 8118:282–299, 2013.

[15] C. Scholl and S. McRoy. Using gestures to resolve lexical ambiguity in storytelling with humanoid robots. *Dialogue Discourse*, 10:20–33, 02 2019.

[16] M. Speicher, B. D. Hall, and M. Nebeling. What is mixed reality? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 05 2019.

[17] A. E. Uva, M. Fiorentino, V. M. Manghisi, A. Boccaccio, S. Debernardis, M. Gattullo, and G. Monno. A user-centered framework for designing midair gesture interfaces. *IEEE Transactions on Human-Machine Systems*, 49:421–429, 10 2019.

[18] D. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9:1–20, 01 2010.

[19] A. S. Williams and F. R. Ortega. Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. *Proceedings of the ACM on Human-Computer Interaction*, 4:1–21, 11 2020.

[20] W. Xu, H.-N. Liang, A. He, and Z. Wang. Pointing and selection methods for text entry in augmented reality head mounted displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 279–288, 2019.