OBJECT MANIPULATION USING SPEECH INTERACTION TO SUPPORT PEOPLE WITH PHYSICAL IMPAIRMENTS

By

Farkhandah Aziz Komal

This dissertation is submitted in partial fulfilment
of the requirement for the degree of Doctor of Philosophy

BIRMINGHAM CITY
University

Digital Media Technology Lab
School of Computing and Digital Technology
Birmingham City University, UK

Dedications

To my beloved father Abdul Aziz (late) and my mother Bushra Khanum for their endless support and sacrifices.

ABSTRACT

Speech input presents an alternative interaction approach that can make creative visual design more inclusive for people with physical impairments. Three core studies were conducted to explore this area further. The first focused on object positioning using three speech-controlled approaches: Speed Control; Location Guides; and Positional Guides. An evaluation with 25 non-disabled participants highlighted a preference for Location Guides which was found to be more usable, efficient, and accurate in positioning objects. A follow-up study with 6 participants who have physical impairments was conducted and validated that Location Guides could support people with physical impairments in accurately positioning digital assets.

A second study investigated object resizing via a prototype that enabled users to alter the size of digital assets using voice-controlled transformation handles. The system was evaluated with 12 non-disabled participants and was found to provide an effective approach for resizing objects, although usability issues were identified. To address these challenges, three new voice-controlled object resizing techniques were developed: NoSnap; UserSnap; and AutoSnap. These approaches were evaluated with 25 participants who have physical impairments where results found that AutoSnap was more efficient, accurate, and usable for resizing assets.

The final study explored the rotation of digital objects via speech input – an elicitation study was conducted with 12 participants having physical impairments to capture rotation voice commands. Findings informed the design of a speech-based prototype for object rotation that was evaluated with 12 participants with physical impairments who were able to successfully rotate digital assets, although some interaction challenges were highlighted. This led to the development of three speech interaction techniques for object rotation: Baseline Rotation; Fixed Jumps; and Animation Rotation. These were evaluated with 25 participants with physical impairments where results found Animation to be more usable, efficient, and accurate than the other approaches. The contributions presented in this thesis highlight that speech input is a viable alternative method that can support people with physical impairments in effectively manipulating digital assets within a creative visual design context.

ACKNOWLEDGEMENTS

I would like to acknowledge and express my warmest thanks to my Director of Studies Dr Chris Creed who provided this opportunity for conducting research under his supervision. His guidance, advice, and continuous encouragement made it possible to complete this research and PhD thesis. I would also like to thank other members of my supervision team Dr Maite Frutos Pasucal and Dr Ian Williams for their valuable suggestions and comments during my PhD. I would like to further add that the support from my supervision team during the pandemic is highly commendable as the original research plan was switched due to COVID-19 circumstances. I am thankful to Dr Sayan Sarcar, and Dr Arthur Theil for their support and feedback during the last year of my PhD. And of course, I am very grateful to all volunteer participants who took part in my research studies as without their support and feedback this research work would not have been possible. I would also like to thank my colleagues and other members of the DMT lab for their advice and support during this journey. I wish to acknowledge Birmingham City University for providing me with a doctoral scholarship and this opportunity to conduct the PhD research.

I would like to thank my parents and my family for their unconditional love, support, and sacrifices they made to fulfil my dream of getting a PhD. I want to dedicate this PhD to my father (late) who wished to see me becoming a ‘Doctor’, I wish you were here to celebrate with me and must have been very proud of me. I would also dedicate this PhD to my mother who is a very strong woman as she continuously encouraged me in achieving this goal despite herself fighting with cancer. I feel sorry for not being able to be with my mother during her cancer treatment but overwhelmed and extremely grateful to her as she did not let me lose hope and continuously pushed to fulfil my dream. I would like to thank both my elder brothers Nauman and Luqman who always inspired me and kept me motivated towards achieving this goal. I want to express a special gratitude to my husband Faizan Ur Rehman for always being there for me in difficult times and for all the love and support throughout this journey. I am thankful to all friends and people who supported me in my struggle in any possible way and made a positive impact on my life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Motivation

The World Health Organisation (WHO) estimates that approximately one billion across the world experience some form of impairment, with around 14.6 million people in the UK affected by disability (Scope, 2023; World Health Organization, 2023). Rapid advancements in technology have led to digital systems becoming a pervasive and essential element of common daily activities such as online banking (Harris et al., 2016), communication with others (Carr and Hayes, 2015), education (Scherer and Tondeur, 2019), business (Mohapatra, 2021), entertainment and gaming (Allison et al., 2018; Reeves et al., 2021; Zargham et al., 2022). However, disabled people can experience significant challenges when attempting to use different technologies thus contributing towards a digital divide where people are excluded from the use of mainstream platforms (Goggin, 2017; Macdonald and Clayton, 2013; Raja, 2016). For instance, people with physical impairments may face numerous barriers when required to use traditional input devices such as a mouse, keyboard, or touch to operate systems (Koester et al., 2013; Trewin et al., 2013). For example, people with Parkinson's disease typically experience tremor in their upper body and may find it challenging to select small targets (e.g., desktop icons, small items in menu bars, etc.) within an interface via mouse input (Keates and Trewin, 2005). Similarly, other conditions such as motor neurone disease, spinal cord injuries, and cerebral palsy can result in issues with muscle control thus contributing towards slow, tedious, and frustrating interactions when using a mouse and keyboard (Gibbons et al., 2013; Mattar et al., 2015; Pousada et al., 2014). To address these accessibility challenges a range of different assistive technologies are available that can support users with controlling digital applications. Common examples include switch-based input (Wilcox et al., 1999), alternative mice and keyboards (Mauri et al., 2006; Szczechowicz and Mazurek, 2018), sip and puff mechanisms (Mougharbel et al., 2013), eye gaze tracking (Majaranta and Bulling, 2014), and brain-computer interfaces (Lazarou et al., 2018). Recent acknowledgement of the importance of inclusive design and creating accessible user experiences for all users has also prompted a shift in the accessibility options available within major operating platforms such as Windows and Macintosh (Halsey, 2015; Koch, 2017). For example, both operating systems provide accessibility features such as altering visual settings, mouse keys, sticky keys, and setting up external assistive tools (e.g., eye trackers) which can support people with physical impairments. Visual settings typically support the adjustment of text and mouse pointer sizes to make target selection easier for people who may experience challenges with motor control. Mouse keys utilise arrow keys on a keyboard (up, down, right, and left) to move the mouse cursor and eliminate the requirement for a mouse, thus assisting users who may have issues with gripping, moving or clicking mouse controls. Sticky keys minimise the use of combination keys (Shift, Alt, Ctrl) during interactions and support users who cannot easily perform these keyboard combinations.

Speech input is another interaction approach that holds significant potential to support people with physical impairments (where their speech in not impaired) and has received increasing interest as an assistive approach in recent years (Cave and Bloch, 2021; Pradhan et al., 2018). In particular, it can enable users to control digital applications without the need to rely on traditional inputs such as a mouse and keyboard (Rebman et al., 2003; Hu et al., 2011; Clark et al., 2019). The rapid growth and success of commercial conversational assistants such as Siri (Adamopoulou and Moussiades, 2020), Alexa (López et al., 2017), and Google Assistant (Dalton et al., 2018) have highlighted the opportunities for voice-controlled systems to present new opportunities to interact with systems (Furini et al. 2020; Göbl et al., 2021; Sotolongo and Copulsky, 2018). Speech interaction has also been utilised to support the production and editing of documents (Kumar et al., 2012; Leijten and Van Waes, 2005; Sears et al., 2003) web browsing and navigation (Sengupta et al., 2018; Cambre et al., 2021), gaming and entertainment (Allison et al., 2018; Rao et al., 2018; Zargham et al., 2022), writing code (Paudyal et al., 2020), and producing creative design work (Kim et al., 2019). Moreover, researchers have specifically investigated the potential of voice input as an accessibility approach within software applications for controlling mouse cursors and performing target selection. For example, studies have explored concepts such as voice cursors (Sears and Karimullah, 2002; Mihara et al., 2005), the selection of interface targets via speech (Dai et al., 2005; Zhu et al., 2009), non-speech sounds to control systems (Harada et al., 2009), and the use of voice in conjunction with other input methods as part of innovative multimodal applications for controlling interfaces (Beelders and Blignaut, 2011; Ronzhin and Karpov, 2005; Srinivasan et al., 2019).

Whilst voice interaction has received increasing interest from researchers to support the development of more inclusive applications, there are areas where there has still been limited work to date. For instance, there is currently a lack of work around how speech interaction can support creative visual design activities (Harada et al., 2007; Van der Kamp and Sundstedt, 2011), ranging from simple object manipulation (e.g., altering the size and dimensions of images and objects) through to more complex creative tasks (e.g., designing application interfaces to a professional standard). This is a crucial area as mainstream creative software (e.g., Photoshop, Illustrator, XD, Figma, etc.) are mainly designed to be operated by only using traditional input methods such as a mouse, keyboard, and touch for producing creative visual design work which cause significant challenges for people with physical impairments. In the context of this thesis, creative visual design refers to designing user interfaces through manipulating digital assets such as images, shapes, icons, and typography, as well as interactive prototyping within a two-dimensional environment.

In particular, digital assets and objects within mainstream applications can easily be manipulated (moved, resized, and rotated) through different controls that can be effectively operated via a mouse or keyboard. For instance, object positioning is a fundamental object manipulation activity which can be performed by either dragging an object via mouse movements or by using arrow keys present on a keyboard. Similarly, object resizing is generally performed through dragging transformation handles attached to an object, while object rotation is also achieved by manipulating transformation handles in clockwise and anticlockwise directions using a mouse. These are fundamental interaction tasks that are essential to support creative design activities, although it is unknown how these common object manipulation activities can be performed efficiently and effectively using speech input. The lack of work on voice as an alternative method in this domain can make it particularly challenging for users who are reliant on speech as an assistive tool to complete even simple creative tasks that many take for granted (e.g., adapting and incorporating basic images and shapes into presentations, reports, or assignments). This can also include a spectrum of users ranging from students who want to develop careers in creative design to professional creative designers who have acquired physical impairment later in life. This is therefore a crucial research area that requires further investigation to explore the extent to which voice control can support users who have physical impairments with basic and fundamental creative visual design activities.

1.2 Research Questions

To address the limited research in this area, this thesis explores new voice-controlled interaction approaches associated with three key object manipulation activities – specifically, the positioning of objects around a design canvas, techniques for supporting the resizing of objects (using common methods such as snapping – Baudisch et al. (2005) and Felice et al. (2016)), and approaches for manipulating the orientation of design assets (i.e., techniques for rotating objects). This is accomplished through investigating the following research questions:

RQ1: How can voice interaction support the positioning of digital objects around a design canvas?

Initial research has investigated positioning of digital objects via simple voice commands such as “left”, “right”, “up”, and “down” (Karimullah and Sears, 2002; Zhu et al., 2010), although this does not fully support users in being able to rapidly move digital assets to specific areas on a canvas with pixel-level precision. Previous work in other related areas (e.g., cursor control via speech) has explored the use of labels located within an interface to support cursor positioning (Dai et al., 2005; Zhu et al., 2009) and is an approach which may hold potential more widely to support efficient positioning of objects. Mainstream design applications also use a range of common features to help facilitate the layout of objects (e.g., alignment guides), although it is unclear whether methods such as these can be adapted effectively for voice control. Further work is therefore required to investigate new techniques that can facilitate the positioning of objects beyond the use of simple navigational speech commands.

RQ2. How can voice interaction facilitate the efficient resizing of digital objects?

Similar to object positioning, there has been some initial work investigating basic voice commands to control the size of digital assets (e.g., using commands such as “larger” and “bigger” – Kim et al. (2019)). However, there has been limited work investigating whether common features used in mainstream design applications that support the resizing of objects (typically operated via a mouse) can be effectively tailored for voice control and thus provide interaction benefits. For instance, designers often use features such as ‘snapping’ via a mouse to assist with resizing objects in relation to other digital assets (located on a design canvas) to help facilitate unity and consistency within a design. The extent to which features such as these can be adapted for speech interaction and whether they can enhance creative workflows for designers with physical impairments is an underexplored area.

RQ3. How can voice interaction support rotation transformations of digital objects?

No research studies to date have investigated how digital assets can be rotated effectively via speech interaction. It therefore remains unclear which types of voice commands designers may find most intuitive to use when performing rotation transformations, as well as the interaction challenges that might exist when utilising speech input in this context. Mainstream creative applications include features that support the rotation of objects via a mouse using transformation handles and incremental jumps in rotation angles (i.e., typically when holding the shift key) to support consistency in designs. However, it is not currently clear how object rotation can be facilitated via voice control – further research is therefore required to investigate new speech-controlled techniques and how designers with physical impairments (affecting upper limbs) perceive the usability of different rotation methods.

1.3 Thesis Contributions

The primary contribution of this thesis is a deeper understanding of how speech interaction can facilitate fundamental digital object transformations (i.e., positioning, resizing, and rotation) to support the creative workflow of designers with physical impairments. The key contributions are detailed below:

- The development of three voice-controlled interaction techniques that support the positioning of digital objects – Speed Control (using transformation speed and simple direction commands such as “left”, “right”, etc.), Location Guides (involving the use of grid-based positional labels), and Positional Guides (alignment guidelines to support object placement) [Chapter 3].

- The first empirical investigation into how these positioning methods can assist with placing objects into specific locations within a design canvas. Results found that non-disabled users preferred Location Guides and that it was more efficient and usable than the other two methods. A further follow-up study validated that users with physical impairments could efficiently use Location Guides to successfully complete standard object positioning tasks [Chapter 3].
- The development of three voice interaction approaches to support the resizing of creative assets, including two techniques tailored around snapping features commonly used in mainstream design applications (i.e., UserSnap and AutoSnap) [Chapter 4].
- The first empirical investigation to explore the potential of voice-controlled resizing and snapping approaches to support people who have physical impairments with object resize transformations. Results found that AutoSnap (automated object snapping to guidelines) was more efficient and usable than NoSnap and UserSnap, and that users with physical impairments had a preference for AutoSnap [Chapter 4].
- The first elicitation study investigating the different types of voice commands that people with physical impairments would utilise to facilitate the rotation of digital assets within a creative context [Chapter 5].
- The development of three different speech interaction methods to support rotation of creative objects – Baseline Rotation (using simple commands i.e., “left/right”), Fixed-Jumps Rotation (incremental jumps in rotation), and Animation rotation (continuous animated rotation) [Chapter 5].
- The first empirical investigation into the potential of speech interaction to support people who have physical impairments with object rotation transformations. Results found that the Animation rotation approach was more efficient, accurate, and usable than Baseline and Fixed-Jumps, and that users with physical impairments have a preference for the Animation method to support rotation of digital assets [Chapter 5].

1.4 Thesis Structure

Chapter 2 highlights different types of physical disabilities, along with a summary of assistive tools commonly used to support users with physical impairments. Speech recognition technology is then introduced with a review of different applications of speech interaction and research exploring the use of voice input for controlling user interfaces. The chapter concludes with a review of research investigating speech interaction in the context of digital creative design work.

Chapter 3 presents the first research studies conducted as part of the thesis focused on investigating object positioning using speech interaction approaches to support people with physical impairments. Three novel voice-controlled object positioning approaches are introduced (Speed Control, Location Guides, and Positional Guides), as well as two different user studies evaluating each technique.

Chapter 4 highlights the next research studies undertaken which focus on exploring object resizing via speech input. An initial exploratory study investigating object resizing using voice control is detailed, including key interaction challenges that were observed during user evaluations (e.g., accurately estimating correct transformation sizes). Three new speech-controlled object resizing approaches are then presented (NoSnap, UserSnap, and AutoSnap), along with details of an evaluation with physically impaired participants to investigate the efficacy of each approach.

Chapter 5 presents the final research studies conducted that investigate different speech interaction approaches for object rotation. An initial elicitation study is detailed where the use of frequently used commands for rotating objects are identified, followed by an exploratory study where these commands were evaluated in an interactive prototype with participants having physical impairments. Three new voice-controlled object rotation techniques (Baseline, Fixed-Jumps, and Animation) are then introduced (informed by results from the previous exploratory study), as well as an evaluation examining the performance of each interaction technique with users who have physical impairments.

Chapter 6 concludes the thesis with main contributions based on findings from chapters 3 to 5, as well as a discussion of limitations associated with the research conducted and recommendations for important future directions in this area.

1.5 Published Work

The following papers have been published as part of this thesis.

- Komal Aziz, F. (2021) Inclusive Visual Design to Support People with Physical Impairments. In: ACM *Companion Publication of the 2021 ACM Designing Interactive Systems Conference (Doctoral Consortium)*, pp. 20-22. **[Core A Ranking]**

  This doctoral consortium paper highlights initial work exploring novel speech interaction techniques to support people with physical impairments with creative design work. The importance of exploring novel speech interaction approaches for three key object manipulation activities (object positioning, resizing and rotation) is highlighted and discussed in this paper.

- Aziz, F., Creed, C., Frutos-Pascual, M. and Williams, I. (2021) Inclusive Voice Interaction Techniques for Creative Object Positioning. In: ACM *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 461-469. **[Core B Ranking]**

  This paper explores three novel speech interaction techniques (Speed Control, Location Guides, and Positional Guides) for positioning objects around a design canvas. Results from an initial user evaluation with non-disabled users found Location Guides to be more efficient and usable than other the other two approaches. A follow-up study with users who have physical impairments also confirmed these results through demonstrating that they were able to successfully complete object positioning tasks using the Location Guides approach. This work is presented in Chapter 3.

- Aziz, F., Creed, C., Sarcar, S., Frutos-Pascual, M. and Williams, I. (2022) Voice Snapping: Inclusive Speech Interaction Techniques for Creative Object Manipulation. In: ACM *Designing Interactive Systems Conference*, pp. 1486-1496. **[Core A Ranking] – (Best Paper Award: Diversity and Inclusion)**

  This paper presents three voice controlled interaction approaches (NoSnap, UserSnap, and AutoSnap) for resizing objects around a design canvas. A user evaluation was conducted with 25 participants who have physical impairments with results highlight that the AutoSnap approach was more efficient and usable than the other two approaches. This work is presented in Chapter 4.

- Aziz, F., Creed, C., Frutos-Pascual, M. and Williams, I. Inclusive Speech Interaction Techniques for Creative Object Rotation. *International Journal of Human-Computer Interaction (IJHCI)*. **[SJR Q1 (Human Factors and Ergonomics) – Impact Factor: 4.920] – In Submission**

  This paper presents three user studies conducted around object rotation via speech interaction. An initial elicitation study was conducted with 12 participants with physical impairments to identify the use of speech commands for rotating objects. This elicitation study informed an exploratory study design evaluated with 12 different users who also have physical impairments highlighting participants' experiences in using voice interaction for object rotation tasks. The findings from this study shaped the design of a final user study with participants who have physical impairments, evaluating three speech controlled object rotation techniques (Baseline Rotation, Fixed-Jumps, and Animation Rotation). Results found the Animation Rotation approach to be more efficient and usable than other two approaches. This work is presented in Chapter 5.

2 BACKGROUND AND RELATED WORK

This chapter initially introduces disability and different forms of physical impairments, followed by an overview and critical reflection of research into assistive tools that people with physical impairments typically utilise for controlling software applications. Speech recognition technology is then presented as another alternative approach for interacting with digital systems, along with a detailed literature review exploring different speech interaction applications and techniques for controlling the user interfaces. Research examining the use of speech input in creative visual design applications is also reviewed, alongside the identification of research gaps within this specific domain.

2.1 Disability and Physical Impairment

Disability is a multifaceted term and is defined as a condition which affects one or more aspects of an individual's life including health conditions, body functioning and structure, social interaction, and many others (Leonardi et al., 2006). Previous work has highlighted how disability can result in significant challenges for a person to participate fully in society and in experiencing equal life opportunities (Fuhrer, 1996; Law et al., 1998). Similarly, while rapid advancements in computer technology have created many technical opportunities, disabled people are often excluded from using these platforms (Sears et al., 2009). A variety of disabilities (e.g., physical, visual, learning, hearing and speech impairments, etc.) have been discussed in the literature in terms of the issues that can be presented in using existing technologies (Burgstahler, 2002; Emiliani and Stephanidis, 2005; Young and Mihailidis, 2010; Ali et al., 2011; Timmer et al., 2015; Brady et al., 2013; Creed et al., 2014). Whilst there are a wide range of impairments that people can experience, the core focus of this thesis is on physical impairments that primarily affect upper body limbs such as hands, arms, or shoulders that present challenges in using traditional input devices (such as a mouse and keyboard). According to Equality Act 2010, physical disability or physical impairment is defined as a condition which limits a person's mobility, physical functioning, or dexterity in a manner that has a substantial or long-term adverse impact on their ability to undertake normal day-to-day activities (Equality Act, 2023). Physical impairment can either be acquired (i.e., through accidents, illness, serious infections, etc.) or congenital in nature (Pape et al., 2002). Similarly, physical disabilities can be situational (e.g., injuries and damaged limbs caused by accidents) or temporary (e.g., due to overuse of equipment) in nature with conditions ranging from mild, moderate, or severe depending on personal circumstances (Necas, 1996). Moreover, individuals with physical impairment can experience varying symptoms across different situational, temporary, or medical conditions and therefore utilise different management strategies, thus making it difficult to generalise the lived experience of physical disability (Brady et al., 2013).

People with physical impairments may also require tailored assistive support or equipment for day-to-day activities. This equipment can include manual or powered wheelchairs, walking sticks, walking frames, artificial limbs, head wands, and mouth sticks (Rai et al., 2019). Manual wheelchairs can be self-operated using large rear wheel handles or via another person pushing a chair, whilst powered wheelchairs are operated using alternative controls (e.g., a joystick type interface attached to the chair). Walking sticks and aids provide balance and stability to people with physical impairments who have limited or reduced walking ability. Moreover, artificial prosthetics that mimic the limb functions (i.e., arms or legs) can help people with limb amputation to perform daily life activities. Head wands (i.e., a stick attached to an individual's head) and mouth sticks (i.e., a stick held in a person's mouth) are also widely used as pointing aids to support people with physical impairments in daily life activities such as writing, drawing, activating switches, and controlling digital systems (e.g., tablets and computer keyboards). Whilst there are a wide range of assistive tools available for people with physical impairments, the need for these tools will largely be dependent on an individual's specific needs and requirements (Bennett et al., 2018; Macik et al., 2014; Wobbrock et al., 2011).

2.2 Types of Physical Impairments

Although physical impairments can affect one or more body limbs, this section discusses further the types of physical impairments that mainly affect the upper body (such as hands, wrist, arms, and shoulders) which can cause significant challenges in controlling digital systems.

2.2.1 Musculoskeletal Disabilities

Physical impairments that mainly affect bones, joints, and muscles are classed as musculoskeletal disorders – for instance, conditions such as repetitive strain injury (RSI) typically affects hands, wrists, fingers, arms, elbows, and shoulders. RSI has also been referred to as a work-related upper limb disorder (WRULD) (Barker, 1995) which can be caused by overuse of technology, pulling or lifting equipment, continuous repetitive movement of hands and arms, twisting movements, or even the use of poorly designed equipment. The affected person can have one or more symptoms that include pain in hands and fingers, tenderness, tingling or burning sensation in hands, stiffness, restricted movements of joints, loss of strength and grip, fatigue, or tiredness (O'Neil et al., 2001). Similarly, tendinitis and tenosynovitis are musculoskeletal disabilities which affect tendons (cords of tissues) that connect body muscles to the bones. The symptoms across these conditions include swelling, inflammation, pain, discomfort in tendons and joints, pinched nerves, and muscles strains. Upper limb amputation such as removal of hands, fingers, or arms caused by trauma (such as an accident) or via surgery (due to different medical conditions) can also result in significant barriers for a person in performing daily life activities (Shahsavari et al., 2020). Arthritis is another musculoskeletal disorder which affects bones, muscles, and joints, and can cause pain, stiffness, restricted movements, and inflammations in joints around hands, knees, and feet (Guccione, 1994). Other examples of musculoskeletal conditions that affect upper body limbs include epicondylitis, bursitis, muscle strains, shoulder impingement syndrome, radial tunnel syndrome, carpal and cubital tunnel syndromes (Aaron et al., 2011; Brukner, 1998; Day et al., 2010; Koester et al., 2005; Pienimäki, 2002).

2.2.2 Neuromusculoskeletal Disabilities

Other physical impairments (referred to as 'neuromusculoskeletal disabilities') affect nerves and brain cells which control muscle movements. For instance, amyotrophic lateral sclerosis (ALS) or motor neurone disease (MND) induce weakness in muscles resulting in a lack of balance, difficulty in gripping, walking and swallowing (Miller et al., 2012). Similarly, spinal muscular atrophy (SMA) is a neurological condition in which a person is not able to move their muscles due to loss of nerve cells present in the brain and spinal cord (Kolb and Kissel, 2011). A person with SMA can have several symptoms including weakness in arms and legs muscles, twitching or shaking muscles, painful muscles in bones and joints, and difficulty in walking and sitting up. Cerebral palsy also affects body movements and co-ordination between brain functions. The symptoms and signs of cerebral palsy vary across different individuals (ranging from mild to severe) – for instance, some people have difficulty in walking and sitting, while others may have severe mobility impairments (Vitrikas et al., 2020). Parkinson's disease is another neurological degenerative disorder that affects brain cells and the nervous system. Physical symptoms include uncontrolled movements, shaking, lack of balance, stiffness, and pain in joints and muscles (Ashour and Jankovic, 2006). Muscular dystrophy is also a neuromuscular disease that is caused by mutation in genes that control the healthy structure of body muscles, resulting in shoulders, upper arms, and leg muscles weakening and shrinking over time (Dalkilic and Kunkel, 2003). Multiple sclerosis is another neurological disorder that affects the brain and spinal cord potentially resulting in limited arm and leg movements, challenges with balance, and a lack of sensation (Alusi et al., 2001). Common symptoms also include fatigue, loss of motor control, and numbness in hands and fingers. Spinal cord injury is also a neuromusculoskeletal disorder that affects the sending and receiving signals from the brain to the rest of the body. This can result in temporary or permanent changes in body functions, movements, and sensations – as well as loss of sensation in hands and fingers, lack of balance, difficulty in walking, back pain, and pressure in the head and neck (Wirz et al., 2010).

2.3 Assistive Technologies for Physical Impairments

Assistive technology is commonly defined as an item, piece of equipment or a system that is used to maintain, increase, or improve functional capabilities of people with disabilities (Matter et al., 2017). Assistive technologies support disabled people with various day-to-day activities such as mobility, communication, and use of technology which can help them live healthy, productive, and independent lives (Scherer, 2002). These can include 'low-tech' assistive tools (e.g., a walking stick, manual wheelchair, etc.) and high-tech solutions such as electronic devices and software applications (Reichle, 2011). There are a large number of assistive technology devices available in the market to assist people with mild or severe disabilities (Dove, 2012). The following subsections present a range of examples of assistive tools that are utilised by people with physical impairments to support their control of digital applications.

2.3.1 Alternative Mouse and Keyboards

Various forms of ergonomic mouse devices (e.g., vertical mouse, trackballs, finger controlled mouse, and trackpads) are used by people with physical impairments to control software (Mauri et al., 2006). These alternative mouse devices can reduce the number of repeated and double clicks performed by users, limit twisting motions of hands and wrist, and can be easier to operate compared with moving a mouse at the desk surface (Wobbrock and Myers, 2006). Vertical mouse is a device that enables users to control the mouse cursor by having a more natural wrist position and can support in avoiding lots of repetitive movements that can assist in reducing wrist pain and discomfort (Radwan et al., 2018). A trackball is also an alternative mouse which remains stationary while the mouse cursor is controlled through an individual manipulating a physical ball using their thumb, fingers, or palm. Similarly, a finger-controlled mouse is a mini handheld trackball operated by thumb control and index finger which can help users in avoiding damage to their wrist and hand muscles, thus reducing the risk of conditions such as repetitive strain injury and arthritis (Jang et al., 2010). A trackpad or touchpad is used to control a mouse cursor by sliding fingers across the touchpad surface which allow the wrist to rest in a natural position since only one or two fingers are needed to operate it (Hertzum and Hornbæk, 2010). Several ergonomic keyboards are also available such as those with split and tented designs, negative slope, and keyboards where key positions have been altered. These keyboards can support movements of upper limbs (i.e., hands, wrist, arms), relieve burden on shoulders and neck, and also help users to avoid accidental keystrokes (Szczechowicz and Mazurek, 2018). Moreover, a virtual keyboard can be used as an alternative mechanism when operated using other input devices such as an eye tracker or a head pointer (Gizatdinova et al., 2018).

Although alternative mouse and keyboard products are helpful in supporting physically impaired users in controlling digital applications, they can also present some unique challenges. Research has shown that prolonged interaction via mouse and keyboards negatively impact the forearm muscles, cause fatigue, and can also lead to other types of injuries such as Carpal Tunnel syndrome and cervical spondylosis (e.g., as a result of users have to adopt a certain sitting postures) (Tiric-Campara, 2014). People who have extensive experience of using a traditional mouse can also find it challenging to adapt to new designs of the alternative mouse (e.g., vertical mouse, trackball, and trackpads) due to the non-mainstream design of these devices (Kim and Ritter, 2015). Similarly, trackballs are not as precise and efficient to use when compared with a traditional mouse (e.g., in the context of fast paced gaming) and may cause strain in an individual's fingers when used over prolonged periods of time (Woods et al., 2003). Touchpads or trackpads on a laptop also tend to be slower in performance and require more sub-movements for controlling a mouse cursor to select targets located at larger distances across user interfaces (due to the smaller size of a touchpad surface) than a traditional mouse (Hertzum and Hornbæk, 2010). Furthermore, adapting to alternative keyboards with altered layouts can present a significant learning curve for users, potentially resulting in increased typing errors when compared with traditional keyboards (Fagarasanu et al., 2005). Some designs of alternative keyboards such as split and tented designs may also have small function keys and lack a dedicated numeric pad which can negatively impact wrist and forearm posture (Rempel et al., 2009). While alternative mouse and keyboard designs can facilitate users in maintaining upper limbs posture, these may also present new challenges affecting overall performance of completing tasks.

2.3.2 Switch Controls

Switch controls are assistive technology devices that are used by people with physical and motor impairments who cannot use a mouse or keyboard to access computers. There are different types of switches that require users to perform push, pull or puff actions to activate a switch (Hackbarth, 2017). Mechanical switches (e.g., Jellybean switches) are commonly operated via a pushing motion on a large button using fingers, hands, arms, feet, or other body movements (Wilcox et al., 1999). These switches are typically connected to computers via a USB connection and can be mapped to specific actions via a software application (e.g., performing a left click and opening a web browser). Users can make use of a single switch for controlling the interface or a combination of multiple switches mapped to different actions (Geytenbeek et al., 2010).

A sip-and-puff switch supports users in interacting with computers by inhaling and exhaling air through a tube or pipe (Grewal et al., 2018). For example, when a user wants to select an interface target, they can either breath in or breath out through a pipe (depending on how the system has been customised). Sip-and-puff switches offer increased accessibility, independence, and control for people with physical impairments by making their interaction possible with computers and humans thus improving the quality of life and empower them. However, sip-and-puff switches also have limitations such as requiring regular cleaning of pipes to avoid the risk of contamination and bacterial infections (Mougharbel et al., 2013).

Furthermore, ‘EarSwitch’ technology enables people with severe physical impairments to perform target selection on user interfaces by contracting their tensor tympani ear muscle (a small muscle that actuates the eardrum) through jaw movements (Röddiger et al., 2022). Although controlling the tensor tympani muscle consistently and accurately can be challenging for users which can lead to unintended actions or difficulties in achieving precise control.

Whilst switch controls can make user interfaces accessible for people with physical and motor impairments, each type of switch requires different levels of force and technique to activate them as discussed, hence the selection of appropriate switch controls can vary across different users depending on the nature of disability. The control of multiple mechanical switches or the combination of different switches can also be difficult to learn for some users and may require additional training (Ali and Cardona-Rivera, 2020). Similarly, accessing interface elements with a single mechanical switch is slow and sometimes painful when scrolling the interface due to repeated movements (Blain et al., 2010). When using a sip-and-puff switch, multiple breath operations may cause respiratory fatigue, while the sip-and-puff pipe also requires regular hygiene maintenance and replacement. In terms of ear switches, many people are unable to easily contract the tensor tympani muscle which can present significant challenges in utilising this method of interaction. Similarly, blocking the ear canal with an ear switch sensor over a prolonged period of time can change the temperature and humidity in the ear canal which may increase the chances of bacterial or fungal infection (Röddiger et al., 2021). In general, whilst switches can provide an accessible and alternative method of input, their overuse can cause fatigue, discomfort, and additional physical limitations and challenges.

2.3.3 Adaptive Controllers

There are several adaptive controllers that are used in gaming environments for example, a joystick controller is an input device that enables users to control the movement and positions of an object in different directions. A joystick consists of a switch knob (i.e., a handle) for controlling user interfaces through different parts of the body such as hands, fingers, palm, chin, head, or foot (Dolan and Henderson, 2017). This versatility enables individuals with a range of physical impairments to use alternative body movements to operate their chosen switch for interacting with user interfaces, hence promoting accessibility and inclusion. A range of other adaptive controllers are available to support people with physical impairments with many innovations in this area being developed within the gaming field. For example, an Xbox controller developed by Evil Controllers (2023) is tailored for use with one hand and provides accessible gaming controls that enable physically impaired people to control gaming interfaces. This controller also provides two input ports at the back of the controller to connect other assistive tools (i.e., switches) that enables players to modify their video game controls so that inputs are more easily reachable, thus enhancing the accessibility of gaming experiences. For instance, users can place the thumb stick of this single-handed controller to any accessible position (via a thumb-stick extension) and can operate it with their mouth, feet, or any other part of their body with mobility. Another adaptive controller is the DualSense device that also enables users to play games using single hand (One-Handed DualSense, 2023). This controller consists of a portable thumb stick which is attached below the controller and can be placed on a user's lap or a table. The Nintendo Switch is video game device consisting of two joy-cons (individual units with an array of buttons) attached to a joy-con grip (Ramolete et al., 2020). The Nintendo Switch can therefore be used as a single handed-controller by separating the joy-cons from the main grip and attaching to a 3D printed adapter (which helps in placing both joy-cons next to each other) making all gaming buttons accessible using one hand. Furthermore, controllers supporting two-handed use (e.g., Xbox 360, Xbox Elite, DualSenseEdge, and Razer Wolverine controllers) are also widely used as adaptive controllers through modifying the video game controls to be utilised by different parts of the body (Byrne and Sweetser, 2022; Gyory, 2022). The Xbox Adaptive Controller (Xbox, 2023) is another example of a device that can support people with physical impairments in controlling gaming experiences. The device consists of two large programmable buttons and nineteen ports which can be connected to a range of other assistive devices such as hand and foot switches, joysticks, and other touch sensitive pads (Garg et al., 2021).

While these adaptive controllers enhance the gaming experience for people with physical impairments, they can also impact a user's ability to efficiently play video games. For instance, the excessive use of joysticks for gaming can potentially cause skin, joint, and muscle problems in relation to repetitive hand movements and joystick control (Weinstein, 2010). Some joysticks also limit the direction of a mouse cursor to only left, right, up, down movements and do not support complex gestures involving diagonal or lateral movements, while these are also not easy to handle operations such as selecting interface elements (Wobbrock et al., 2005). Furthermore, some people may find it difficult to maintain precise control when using joysticks – for example, tremors and involuntary movements may cause inaccurate clicks thus leading to frustration and difficulty in performing tasks accurately. The use of modified single-handed gaming controllers (i.e., DualSense and Nintendo Switch controllers) for prolonged duration can also adversely affect upper limb muscles, especially if individuals have an impairment in their dominant hand (Hassan et al., 2022; Maggiorini et al., 2019). Moreover, one-handed controllers can impact a user's ability to play games competitively against those using two-handed controllers if controller buttons are not easily accessible via the dominant hand. A wider accessibility point relates to gaming controllers typically requiring batteries that need to be frequently changed which can present challenges for people with physical impairments (i.e., the process of changing batteries requires a certain level of dexterity in an individual's hands) (Gonçalves et al., 2021). Additionally, whilst devices such as the Xbox Adaptive controller allows users to attach multiple assistive tools, the process of configuring, managing, and controlling additional input modalities can contribute towards increased cognitive workload (Wentzel et al., 2022). Adaptive gaming controllers can also be expensive in comparison to other interaction modalities such as speech input which is integrated within mainstream operating systems (Martin, 2020). Furthermore, the use of modified gaming controllers can require significant practice to be able to effectively handle gaming controls (Oshita and Ishikawa, 2012). Overall, whilst these devices can facilitate more accessible interactions for people with physical impairments, they can also lead to new challenges associated with input efficiency, performance, and fatigue as discussed.

2.3.4 Eye Gaze Interaction

Eye tracking technology can be particularly useful for people with severe physical impairments to control digital systems using their eyes as the physical control of additional devices is not always required (Galante and Menezes, 2012; Majaranta, 2011; Rotariu et al., 2019). Early research on eye tracking technology typically used standard web cameras to track eye movements for text entry and controlling user interfaces (Hansen et al., 2001; Grauman et al., 2003; Magee et al., 2004). However, there are multiple challenges that impact accurate detection of eye gaze movements when using a webcam such as varying light conditions, reflection and shadows (Goldberg and Wichansky, 2003). Eye blink switches also allow people with physical and motor impairments to control various software applications including selecting interface elements, navigation and web browsing, and playing games by blinking their eyes. An eye blink switch can detect eye movements using a web camera, although certain factors such as a low quality camera, presence of glare and changes in head position can limit the accuracy of blink detection to accurately perform an action (Udayashankar et al., 2012; Singh and Singh, 2018; Tanwear et al., 2022). In contrast, infrared-based eye trackers tend to be more accurate through tracking the reflection of infrared light (emitted from the tracker) off the cornea (transparent front part of the eye) and iris (colored part of the eye) (Majaranta and Bulling, 2014). Infrared-based eye trackers are more widely available as commercial devices (over those using web cameras) due to their accuracy, reliability, and compatibility with different platforms (e.g., the Tobii suite of tools – Kasprowski, 2022).

A large number of studies utilising this technology have been conducted with a particular focus on eye typing and target selection (Duchowski, 2002; Majaranta and Räihä, 2002). A key interaction challenge with this technology is the well-known Midas touch problem where accidental selections can be trigged when users shift their gaze across user interfaces (Jacob and Karn, 2003). To address this challenge, research has focused on approaches such as the use of dwell time in which a user's gaze must remain fixated on a target object for a specific duration to trigger a selection (Majaranta et al., 2009; Nayyar et al., 2017; Huckauf and Urbina, 2011). Studies have also explored alternative dwell-free methods such as Dasher which is a predictive text entry method where a stream of alphabet letters continuously flow towards a user's gaze direction to select their desired character (Rough et al., 2014). Whilst this has been shown to support faster typing speeds, it also requires significant practice and cognitive effort to achieve optimal typing speeds (Benligiray et al., 2019). The use of gaze gestures is another approach which utilises intentional gaze movements in specific directions to support object selection, thus helping to avoid accidental clicks (Heikkilä, 2013; Schenk et al., 2017; Wibirama et al., 2020). However, this approach can also lead to fatigue and increased cognitive load through the repeated actions required in consciously controlling gaze movements (Slobodenyuk, 2016). Studies have also explored the use of multimodal approaches to avoid accidental clicks through combining eye gaze interaction with mechanical switches (Creed et al., 2020), head gestures (Feng et al., 2021), foot pedals (Klamka et al., 2015), and speech input (Sengupta et al., 2018), although this approach can add technical complexity as users have to control multiple assistive devices (Silva et al., 2015).

Another key problem with eye gaze interaction is the selection of small interface targets which can be challenging due to limitations around the accuracy of eye trackers. Researchers have explored different approaches to address this issue – for instance, Skovsgaard et al.

(2010) investigated zooming approaches such as two-step zooming (with 4x magnification) and three-step zooming (with 8x magnification) with variable dwell times and found three-step zooming (with shorter dwell time) was faster and accurate than two-step zooming for small target selection. Similarly, Choi et al. (2020) investigated the use of a bubble gaze cursor (i.e., a circular cursor) which expands targets where a user's gaze is currently fixed, thus making smaller objects easier and more accurate to select over other zooming techniques. Lutteroth et al. (2015) presented a hyperlink selection approach for web navigation by assigning distinct colours to each hyperlink present within a webpage, while larger corresponding buttons (in the same colours as hyperlinks) were displayed in the browser sidebar to facilitate object selection via eye gaze. The authors found that their approach was more accurate for small target selection in comparison to a mouse and gaze only click alternative. However, this approach requires additional steps such as navigating gaze into the sidebar to confirm object selection which could be tedious and frustrating for users. A further challenge with eye tracking technology is calibration drift where the tracking system can lose its initial calibration (e.g., due to factors such as different lighting conditions, changes in pupil size, and user's head movements), resulting in inaccurate tracking experiences (Müller et al., 2019; Sugano and Bulling, 2015). Furthermore, eye tracking technology may not offer sufficient customisation options (e.g., altering dwell time speeds or tracking of one eye instead of two) for individuals across a diverse range of physical impairments, hence limiting the accessibility of this technology (Duchowski, 2018; Hyrskykari et al., 2012; Vessoyan et al., 2023). The high cost of eye tracking hardware and corresponding software can also present a significant barrier in accessing this technology, hence preventing wider adoption of eye gaze interaction (Titz et al., 2018).

2.3.5 Brain Computer Interfaces (BCIs)

Brain Computer Interfaces (BCIs) are another technology that hold potential to support people with severe physical impairments (who are unable to speak or use their limbs) to interact with computers and communicate with other people. BCIs translate brain signals into commands that are received by an external application which can be used to control digital applications or smart home systems (Lazarou et al., 2018). There are both non-invasive and invasive forms of BCIs – non-invasive BCIs are based on EEG (i.e., Electroencephalography) and receive brain signals by attaching multiple electrodes externally to the scalp (Cincotti et al., 2008). Commercial non-invasive BCIs are typically more affordable than invasive methods and do not require the intervention of medical professionals (Kim et al., 2014). However, signal transfer rate in non-invasive BCIs is low and signals are often distorted due to the distance between neurons in the skull and externally attached electrodes (Kalagi et al., 2017). Invasive BCIs require surgical implantation of a device to ensure direct contact with the brain's surface. Whilst the quality of brain signals to the system is much stronger in this scenario, this approach can also damage neurons present inside the brain due to the required surgical procedure (Becedas, 2012).

Studies have looked at detecting, interpreting and translating imagined speech into computer generated output (Schultz et al., 2017). Brumberg et al. (2018) utilised this approach to support people with severe motor impairments in manipulating the positions of circular objects using imaginary movements of their hands and foot. Sereshkeh et al. (2018) explored the use of imagined speech to convey an answer to different questions by mentally repeating the phrases 'yes' and 'no' which they suggest could be useful in a number of scenarios (e.g., calling caregiver, activating an assistive device, and playing music). Furthermore, Lee et al. (2022) utilised imagined speech to support smart home control within a virtual world. They evaluated the system with non-disabled participants using the imagined speech of thirteen words (i.e., hello, help me, ambulance, clock, light, pain, stop, thank you, toilet, TV, water, and yes). However, despite tremendous amount of training the BCI system struggled in accurately determining the thoughts of non-disabled participants. The accurate signal acquisition for imagined speech systems could even be more challenging when participants with motor and physical impairments are involved as this can require several hours of training and increased cognitive effort for them.

Current BCIs still have limited capabilities and provide mainly visual feedback or basic communication which is relatively slow and often inaccurate. For instance, electrical signals are challenging to properly detect in the case of non-invasive BCIs which can cause difficulty in effectively interacting with a system. BCIs are also susceptible to noise and interference from internal and external sources (e.g., muscle activity or environmental factors) which can affect the reliability and overall performance of the BCI systems (Yuan et al., 2021). Whilst there remain accuracy challenges with this technology, BCIs still hold the potential to be a viable interaction mode in future scenarios for a variety of applications (e.g., efficiently controlling other assistive devices, wheelchairs, and robotic arms) if technological progress can be made around supporting better signal acquisition, designing more convenient hardware, and through increasing general reliability of the technology (Gu et al., 2021), However, this still remains an open research area where further work is required to enhance the viability and efficacy of the technology. Since BCI technology is mainly used by a relatively small user population, there has not yet been a significant incentive for commercial organisations to manufacture and promote their widespread use (Shih et al., 2012). Furthermore, the development of BCIs for people with physical impairments still requires clear validation of their real-life value in terms of efficacy, practicality (including cost-effectiveness), and impact on quality of life (Camargo-Vargas et al., 2021).

2.3.6 Summary

Whilst there are a range of assistive technology tools that can support people with physical impairments in controlling user interfaces, each of these technologies also present different challenges. Alternative mice such as vertical mice, trackballs, finger-controlled mice, and trackpads offer ergonomic support that can reduce strain on the hands and wrists, as well as helping users avoid repetitive movements, wrist pain and discomfort. However, they may require users to adapt to unnatural and non-mainstream designs and can have limitations in precision or performance. Similarly, while different switch controls (i.e., mechanical switches, joysticks, ear switches, and sip-and-puff) offer benefits for individuals with physical impairments, users may find it challenging to learn how to effectively utilise switches to control applications. Similarly, overuse of switches can not only cause fatigue and discomfort but may also lead to other type of impairments impacting muscles, joints, skin, and other health conditions. While adaptive gaming controllers enhance accessibility through customization, they can also lead to increased fatigue, and muscle strain. Other issues include battery replacement, configuration complexity, and the need for additional practice to effectively handle gaming controls. Gaze interaction in another alternative interaction method to enable individuals with physical impairments to control the mouse cursor and navigate interfaces by using eye gaze. Various eye typing and target selection techniques have been used including dwell time, dwell-free methods, gaze gestures, and multimodal gaze interaction. Although, certain challenges such as lighting conditions, reflection, and other challenges including accidental clicks (i.e., Midas touch), and calibration drift impact the overall accuracy of gaze interaction and user experiences in controlling interfaces. Brain computer interfaces also aim to provide accessible experience to people with severe physical and motor impairments using invasive and non-invasive procedures. Invasive BCU involve painful surgical implantation of sensors inside brain while non-invasive BCI are slower in capturing the brain activity. In general, BCIs require extensive training for each individual user while unintentional muscle activity or environmental factors can still impact the accuracy and performance of the system. While these assistive technologies offer valuable options for individuals with physical impairments to interact with software applications. However, they also present unique challenges in terms of adaptation, performance, fatigue, hygiene maintenance, and cost.

Another technology that has received significant attention from researchers to support people with physical impairments is speech recognition and interaction. Speech technology aims to bridge the gap between humans and machines by enabling communication through spoken language. It encompasses a range of technologies, including speech recognition, synthesis, and natural language processing, to enable devices and systems to understand, interpret, and respond to human speech. Speech technology holds significant promise in enhancing communication and interaction for those who face challenges in using traditional means of input (i.e., a mouse and keyboard). In particular, for individuals with physical impairments, speech technology offers a significant opportunity to overcome current interaction barriers and limitations associated with controlling physical assistive devices, thus fostering increased independence in accessing and controlling digital experiences. This technology forms the core focus of this thesis and is reviewed in detail within the following subsections.

2.4 Speech Technology

Human speech is the most common and natural method of interaction that is used between humans to transfer ideas and knowledge (Kumar and Singh, 2019). The use of speech also offers potential advantages for human-computer interaction as it can potentially facilitate a more natural paradigm for controlling and instructing systems (Reitmaier et al., 2022). Previous research has also showed that speech interaction could be faster than using traditional inputs such as a mouse and keyboard and users prefer the use of speech technology over traditional inputs (Cohen and Oviatt,1995). The use of speech recognition in digital applications can present significant opportunities as an assistive tool to support people with physical impairments in accessing different computing platforms (Cave and Bloch, 2021). However, it can also present interaction challenges that may result in user frustration and usability issues which are discussed later in this chapter. Key work associated with this field is reviewed below to highlight the potential of voice interaction to support people with physical impairments whose voice has not been significantly impaired. This literature review was guided by the following questions:

- Which types of speech interaction techniques have been utilised for controlling digital applications?
- Which types of speech interaction techniques have been explored for manipulating digital assets in creative design applications?
- Can existing speech interaction techniques support people with physical impairments in manipulating digital objects in creative design applications?

Search Process

The search process included an initial database search, a manual search of publication venues in the HCI field, and a final database search using an extended list of keywords. The list of keywords used in the search process included ‘accessibility’, ‘accessible applications’, ‘alternative interaction’, ‘interaction techniques’, ‘automatic speech recognition’, ‘speech interaction’, ‘speech interaction techniques’, ‘multimodal interactions’, ‘physical impairment’, ‘physical disabilities’, ‘inclusive design’, ‘creative work’, ‘creative design applications’, ‘designing tools’, ‘interface designing’, ‘designing activities’, and ‘object manipulation’. A search process was initiated using these keywords via digital library services such as Google Scholar, ACM Digital Library, IEEE*X*plore, Scopus, ScienceDirect, ResearchGate, and Springer. All papers and their related data (such as the paper title, list of authors, publication venue, and publication year) were stored using the Mendeley reference manager. The search process resulted in a total of 260 research papers being identified. A secondary search was then conducted using a backward snowballing approach where the list of references from collected papers were examined to find additional relevant literature. This secondary search identified 42 additional research papers, resulting in a total of 302 papers being collected during the selection process. After the search and selection process, the inclusion and exclusion criteria highlighted below were applied based on the analysis of titles, abstracts, and paper conclusions to ensure relevant papers were chosen for literature review.

Inclusion Criteria

- Research papers that focused on different speech interaction techniques (both speech only and speech as a multimodal interaction) in digital applications for object manipulation activities.
- Research papers with the main emphasis on controlling digital applications to support people with physical impairments in object manipulation activities.
- Research papers that focused on design and evaluation of speech interaction approaches via speech only or speech as multimodal approach for creative design work.
- Research papers that focused on speech interaction approaches via speech only or multimodal speech interaction for object manipulations in creative work.
- Research papers that were published in HCI conferences and journals.

Exclusion Criteria

- Research papers related to assistive technologies to support people with physical impairments and speech impairments.
- Visual artistic assistive tools to support non-digital drawing activities.
- All duplicate research papers that were extracted and downloaded from various sources and databases.
- Studies that are prereview papers, newsletters, surveys, and non-English articles.

Papers that did not provide sufficient information through title, abstract, and conclusions, but contained relevant keywords were fully read to consider their relevance to the review. Out of 302 papers, a total of 162 papers were retained and again read fully to further analyse their methodologies to see relevance against the criteria identified. This final process resulted in 75 papers being selected for the review.

2.4.1 Speech Technology Overview

Speech technology can be classified into two major areas – speech recognition and speech synthesis (Schmidbauer et al., 1993). ‘Automatic Speech Recognition’ (ASR), ‘speech-to-text’ or simply ‘speech recognition’ enables machines to understand human speech and convert it into a series of words to perform an action (Benkerzaz et al., 2019). Early research on speech technology was conducted in the 1950s at Bells laboratory where a speech recognition system (‘Audrey’) was developed that could recognise ten English digits when used by a single user via voice input (Meng et al., 2012). Similarly, in the late 1950s, Forgie and Forgie at MIT Lincoln Laboratories created a system which could recognise up to ten vowel sounds when issued by users (Gibbon et al., 1997). Whilst these projects presented important landmarks in developing systems capable of recognising voice input, they still lacked the ability to accurately process continuous speech. During the 1960s to 1970s, several attempts were made to join speech utterances with Reddy (1976) being the first to conduct research on continuous speech recognition, culminating in a system that could continuously recognise up to 100 words (Furui, 2005). Later in 1971, a DARPA funded project built the Harpy speech system which was able to continuously recognise over 1000 words (Francis and Nusbaum, 1999).

During the 1980s, the core focus of research in the field was focused around creating more robust systems capable of recognising fluently spoken strings of connected words (Furui, 2010). With the rapid advancement in dynamic programming and statistical modelling around this time, significant improvements were made in the development of speech recognition applications. In particular, Hidden Markov Models (HMM) became one of the key voice recognition algorithms in the 1980s that were widely used in a majority of research laboratories. Fred Jelinek at IBM created a voice typewriter ‘Tangora’ which was able to handle a vocabulary of 20,000 words (Averbuch et al., 1987). Dragon software was also founded in 1982 and released as a commercial speech recognition application, becoming a popular choice for users over subsequent years (Dragon, 2023).

During the last three decades, voice recognition has been enriched by the power of machine learning and artificial intelligence. Further attempts were made to increase the robustness of speech recognition such as spontaneous speech detection, identification of semantically significant parts, rejecting irrelevant portions in spontaneous commands, and improving speech recognition accuracy (Deng et al., 2010; Ghannay et al., 2018; Munteanu et al., 2006; Silsbee and Bovik, 1996). In 2008, Google launched ‘Voice Search’ for the iPhone and later in 2011 Apple introduced ‘Siri’ – a digital personal assistant for supporting users (Sonix, 2023). In 2011, the W3C Speech API community group and members of Google released the ‘Web Speech API’ which presented new possibilities for accessibility and controlling web interfaces (Web Speech API, 2023).

2.4.2 Speech Interaction Applications and Techniques

In recent years, ASR has been used in a range of applications including conversational or digital assistants (e.g., Siri, Alexa, etc.), dictation software (speech-to-text), web applications, augmented and virtual reality, voice search, health care applications, and many other areas (Honeycutt, 2003; Hu et al., 2011; Johnson et al., 2014; Williams et al., 2020). Speech input can also play an important role in several situations – for example, navigating through multiple menus, performing complex searches, during driving when hands are occupied, or simply when natural language interaction is a user's preference (Corbett and Weber, 2016; Lugano et al., 2017; Robinson et al., 2019).

Voice controlled virtual assistants and devices (e.g., Siri, Cortana, Alexa and Google Home) are a relatively new interaction paradigm that provide conversational interactions in home and personal environments to support a variety of activities (Pradhan et al., 2018). Conversational interfaces facilitate two-way communication by enabling users to interact with software following a principle similar to human-to-human conversation (Klopfenstein et al., 2017). With virtual assistants, a user can perform voice searches to (for example) get updates on the weather forecast and traffic jams (Lopatovska et al., 2019). Similarly, a user can perform a range of other actions such as online shopping, banking, playing music, scheduling appointments, as well as controlling smart devices such as a home's temperature and lights (López et al., 2017; Purington et al., 2017; Vassilev et al., 2020; Ramadan et al., 2021; Çepik et al., 2021). In terms of mobile devices, Corbett and Weber (2016) presented a voice interaction approach to support people with physical impairments in selecting icons, navigating an interface, launching applications, calling others, and sending messages (through assigning numbers to all icons present on the mobile interface that could be selected through vocally stating the relevant number). Chang et al. (2021) explored a voice-based navigation approach inside an online video based on the titles of different segments. For instance, instead of issuing a "skip 20 seconds" command, users could directly state "cup of sugar" (i.e., a segment title) to skip to that specific part of the video. Moreover, Carter et al. (2015) investigated the use speech input for playing video games to support users with selection and navigation controls via voice commands such as "play", "pause", "move" and "show map".

There are also some interaction challenges associated with speech technology (Alsuraihi and Rigas, 2007) – for instance, when people with different accents try to communicate with smart speakers there could be issues with recognition accuracy resulting in users having to repeat commands. Similarly, when using computers or mobile devices in a public place, surrounding sounds within the environment can also impact speech recognition accuracy thus resulting in usability issues. Vocal fatigue and throat discomfort can also present challenges for users when speech commands are used repeatedly over a prolonged period of time (Kambeyanda et al., 1997). Despite these challenges, research has also highlighted that speech interaction is a promising alternative to traditional inputs such as mouse and keyboard which can present benefits in a range of scenarios (Sears et al., 2003). For example, research has explored speech interaction techniques in terms of controlling cursors and selecting targets within interfaces, methods for writing and correcting text, and how it can be combined with other input modalities. An overview of work in these areas is provided below.

2.4.2.1 Speech Cursors

The concept of a “speech cursor” has been explored in previous literature to help users position a mouse pointer in different directions using voice commands. Karimullah and Sears (2002) used six voice commands – “move left”, “move right”, “move up”, “move down”, “stop” and “click” to continuously move and control the mouse cursor when performing target selections within graphical interfaces. In another study, Sears et al. (2003) evaluated a speech-controlled cursor to measure its performance in acquiring both larger and smaller targets over shorter and longer distances. Results found that there were fewer errors and shorter selection times in relation to larger targets, although the authors highlighted cursor speeds were a crucial factor in selecting smaller targets. Similarly, Sporka et al. (2006) conducted a preliminary study to evaluate speech cursors using two different modes, orthogonal and melodic. In orthogonal control, the mouse cursor moved with variable speeds in vertical or horizontal directions. While in melodic control, the mouse pointer moved in the same directions with a fixed movement speed. Results suggested that participants found the orthogonal approach easier to use due to the additional control they had over manipulating cursor speeds. Mihara et al. (2005) investigated the use of multiple ‘ghost’ cursors aligned vertically and horizontally in the form of intersecting lines to help users with precise selection of small targets. These cursors were supported through the use of continuous non-verbal speech commands (i.e., an “Ahhhh” sound) to reach at a specific cursor closer to a target location. The authors reported that this approach enabled users to acquire target positions accurately, although evaluation results also indicated that participants perceived it as slower than a standard mouse cursor as they had to use continuous non-verbal speech commands to navigate through multiple ghost cursors.

Whilst speech cursors have been shown to provide an alternative input method for people with physical impairments, they can also be slower and more tedious to use than a traditional mouse if users have to navigate through a grid of cursors to acquire a target (Zhu et al., 2010). Research has also demonstrated that it is challenging to maintain precise control over a mouse cursor when a speech-based cursor is designed to continuously animate towards a target object (Feng et al., 2005). Therefore, this approach can result in inaccurate clicks being activated (due to a delay in the recognition of voice commands), as well as requiring additional effort in positioning a cursor at a specific target position.

2.4.2.2 Target Selection via Speech

Studies have also explored the concept of grids controlled through voice commands to help users with target selection within graphical interfaces. For instance, Dai et al. (2003) used a 3x3 grid where each cell was assigned numbers from 1 to 9 (from upper-left to lower-right), along with a single cursor located in the middle of each cell (i.e., a total of 9 cursors placed in a 3x3 grid). A user could select a desired target through issuing a corresponding voice command (e.g., “select 5”) to enter into a specific cell located closest to the target position. The authors found this grid solution to be faster for target selection (compared with a one cursor solution), but also demonstrated higher error rates in terms of selecting small targets. Similarly, Zhu et al. (2009) proposed a speech-based 3x3 grid to support the selection of small targets using a single cursor, although they also utilised additional magnification and fine-tuning methods. The magnification feature supported users with zooming into a specific cell recursively, while the fine-tune feature was used to position the grid using simple direction commands (“left”, “right”, “up”, “down”). Results highlighted that using both the magnification and fine-tune features significantly enhanced target selection performance.

Dragon Naturally Speaking software (Dragon, 2023) introduced a similar speech-based grid to provide users with hands-free control for desktop and web applications. In order to activate a grid, users issue a “MouseGrid” command which displays a 3x3 grid over the full screen. Each grid cell consists of numbers from 1 to 9 (from top-left to bottom right). Users can then specify the number of grid cell (e.g., “8”) to position the mouse pointer over the relevant cell. If a target is not located at the exact location of the onscreen cursor (when moving into the cell), users can again issue another numbered command for a smaller

MouseGrid that appears over the target area. Once the mouse cursor reaches the user's desired target position, they can then issue a "left click" command to complete the selection. Overall, whilst speech-based grids have been shown to offer benefits for target selection (compared to traditional input devices), they also typically involve additional steps (such as zooming recursively) for accurate selection of objects which can be tedious and frustrating for users (Zhang et al., 2020).

2.4.2.3 Non-Verbal Speech Input

Previous literature has also explored non-verbal speech for performing certain actions (Harada et al., 2008). For example, the use of an "ahhhhhh" to turn up the volume on a TV set (Igarashi and Hughes, 2001) or the use of vowel sounds (e.g., "a" sound for "up", "e" sound for "right", "i" for down, and "o" for left) for controlling and moving a mouse cursor with consistent speed in a specific direction (Harada et al., 2006). The use of non-verbal speech commands can present interaction benefits through reducing the need for repeated verbal commands. For example, instead of repeating "up", "up", "up" commands, users can issue a continuous 'aaaa' sound to continuously move the mouse cursor in an upward direction. More widely, non-verbal continuous speech has been used to manipulate robotic arms – for instance, an "iy" sound for moving the arm in a left direction and an "aw" sound for moving to the right (House et al., 2009). Studies have also used non-verbal speech for basic manipulations in gaming experiences - for instance, humming and whistling sounds for "left" and "right" actions which have been found beneficial to support people with motor impairments (Sporka et al., 2006). Furthermore, the use of 'blowing' sounds have been investigated to support the selection of objects within digital applications, although usability evaluations found it to be exhausting to use in comparison to a standard joystick (Zielasko et al., 2015). Harada et al. (2011) also utilised non-verbal speech interaction to provide users with a hands-free gaming experience (i.e., through vowel sounds to control movements and directions of game characters) with user evaluations results highlighting that it was preferred by users over mouse and keyboard control. Whilst non-verbal speech can be useful in completing simple tasks, it can be challenging to use vowel sounds for complex operations such as performing creative design work (e.g., efficiently manipulating an object from different directions and simultaneous transformations of multiple objects). Similarly, switching between different non-verbal sounds (i.e., "aaaa", "eeee", "iiii", "oooo" etc.) to continuously perform various different actions can be tedious and frustrating (Pan et al., 2023). Multiple pauses when issuing non-verbal speech commands can also make it difficult to produce longer and smooth curves using a digital drawing pen. Furthermore, it can be challenging to produce smooth brush strokes if there are fluctuations in vowel sound utterances (e.g., caused by vocal fatigue) which can result in commands not accurately mapping with drawing actions (Harada et al., 2007).

2.4.2.4 Multimodal Speech Interaction

Speech interaction has also been explored in multimodal interactions to enable users to interact with digital applications using the combination of multiple input modalities (Turk, 2014). Examples of these multimodal interactions include a combination of speech input with traditional inputs (mouse and keyboard), eye gaze, gestures, and head tracking. Ismail and O'Brien (2008) conducted a preliminary study using a multimodal approach (speech and mouse input) for search and navigation activities associated with digital photos. Results from a user evaluation highlighted that participants had more positive perceptions of the multimodal approach when compared with a traditional mouse only approach. Kane et al. (2011) presented a multimodal system in which touch input was used for selection and speech input for searching while manipulating a map. A user evaluation found that the combination of touch and speech was more effective than touch only input where users had to use repeated tapping actions to find a place on map. Saktheeswaran et al. (2020) also explored the use of voice control in combination with touch input for search and navigation activities on mobile and tablet devices. The authors highlighted enhanced performance and a higher level of user satisfaction with the multimodal interaction when compared to touch or speech-based interaction alone. Williams and Ortega (2020) explored the combination of speech input and gestures as a multimodal approach for manipulating objects in 3D environments. Results from an evaluation highlighted that the combination of speech and gestures performed better than gesture only or speech only inputs. Ronzhin and Karpov (2005) presented a multimodal system based on head tracking and speech input for cursor navigation and target selection tasks. Results from a user evaluation demonstrated that the multimodal system was easier to use and less physically demanding than a traditional keyboard and mouse approach. Moreover, Beelders and Blignaut (2011) explored multimodal interaction for word processing tasks using the combination of eye gaze (for eye typing via onscreen keyboard) and speech input (for text formatting and moving cursor position). The authors found that the multimodal approach (speech and eye gaze) was more efficient and effective in terms of task completion time and accuracy in comparison with eye gaze only input for word processing activities. Research has also investigated multimodal speech interaction approaches (i.e., speech input + eye gaze) to provide users with hands free web browsing experiences and to perform different operations such as web searching, selection of web links, and bookmarking web pages (Heck et al., 2013; Sengupta et al., 2018). Furthermore, speech input as a part of multimodal approach (i.e., in conjunction with eye gaze) has been investigated in coding environments to support people with physical impairments in writing and editing code (Paudyal et al., 2020).

Whilst multimodal interactions offer flexibility and can be efficient to utilise in some scenarios, they also present challenges around how different modalities should be integrated and controlled by users with physical impairments (Song et al., 2012). Multimodal approaches using mouse or touch inputs in combination with speech input has been shown to be physically demanding as people with motor and physical impairments can find it difficult to perform precise hand movements and maintain steady control of a mouse or touch input (Wentzel et al., 2022). The combination of gestures with speech input as a multimodal approach may cause interaction challenges for people with physical impairments due to involuntary body movements, challenges around wearing devices and sensors securely, as well as tiredness and fatigue associated with the repeated use of gestures (Creed et al., 2023). Similarly, the combination of speech input with a head tracker can introduce usability issues due to small involuntary movements potentially resulting in unintended actions being performed (Kong et al., 2019). Furthermore, speech input in combination with eye gaze technology can also present challenges through the need to configure and manage multiple input tools which can lead to the increased likelihood of usability issues (e.g., due to calibration drift and challenges with speech recognition) (Müller et al., 2019). Multimodal approaches have also been shown to require significant additional cognitive workload to operate them effectively due to the need for managing two or more interaction modalities for simple tasks such as navigation and target selection (Turk, 2014). On the other hand, a unimodal interaction approach such as speech-only input offers users a handsfree approach to interact with systems using only their voice (Reitmaier et al., 2022). This approach could potentially be more feasible for people with physical impairments and reduce the need within multimodal interactions for users to control multiple devices (which can lead to additional cognitive load and usability issues).

2.4.3 Speech Interaction for Creative Design Work

Previous literature has explored the potential of speech interaction to support creative design work within multimodal approaches. In early work, Hauptmann (1989) evaluated different interaction approaches (speech only, gesture only, and multimodal approach using a combination of both speech and gesture) for moving and scaling a cube via a limited vocabulary set (i.e., “left” and “right” commands). Overall, results from a user evaluation highlighted that the majority of participants preferred multimodal interaction for manipulation tasks. However, only a small set of manipulation operations were evaluated using a single cube shaped object. Post-study feedback also highlighted that participants encountered significant issues when using speech only interaction due to misrecognition of commands, although it is important to note that early speech recognition systems utilised relatively basic language models which may have contributed towards the higher inaccuracies (Furui, 2010). Moreover, the results from this work were limited to qualitative feedback with no statistical comparison conducted between the different conditions to investigate different performance metrics (e.g., task completion times, accurate positioning, and usability ratings) which can help to validate the reliability of the presented approach. Pausch and Leatherby (1991) also presented a graphical editor (MacDraw) for basic drawing operations using multimodal interactions (i.e., speech input and traditional mouse and keyboard). The selection of drawing tools and basic actions were performed using voice commands such as “arrow”, “rectangle”, “polygon”, “cut”, “paste”, “select all”, and “undo”, while drawing operations were performed using mouse input or a keyboard. The authors compared the use of traditional inputs with a multimodal approach and found the combination of speech input with a mouse to be faster in terms of task completion time than the combined use of a keyboard and speech input. However, their user evaluation involved only novice participants (i.e., in terms of technical experience and design skills), therefore it is unclear whether these results would also apply to a more diverse range of users with different levels of experience. The study also only included the use of a small number of speech commands as the majority of their drawing activities were supported via traditional input. It is therefore not clear whether the findings would be consistent if a wider range of additional voice commands were required to be utilised (e.g., to access variety of drawing features or interface properties. Similarly, Gourdol et al. (1992) also proposed a multimodal interaction technique combining speech with mouse and keyboard inputs within their VoicePaint application. The application supported drawing operations such as selecting paint brushes and sizes, adding shapes, colour selection, and basic manipulations (i.e., moving and scaling objects) via voice input and a mouse, while text entry was performed using a keyboard. However, the authors did not provide specific details around the list of available voice commands used for drawing operations. Furthermore, no user evaluation was conducted thus making it challenging to determine the viability of the application to support users with drawing activities.

Hiyoshi and Shimazu (1994) also presented a multimodal approach to support object positioning where mouse pointing was used to specify a target position for a basic shape and speech input was used to complete movements (e.g., via statements such as “place the object here”). The authors suggested that the combination of mouse input and speech is useful as a multimodal approach for manipulating objects, although no user evaluation was presented to validate the efficiency and reliability of the approach presented. Furthermore, Nishimoto et al. (1995) investigated the use of speech input for drawing basic shapes (i.e., rectangle and circle) and compared its performance with a mouse and keyboard. Users performed basic drawing actions on rectangle and circle shapes using voice commands such as “cut”, “delete”, “paste”, and “undo”. Results from a user evaluation highlighted that speech input performed better than traditional inputs and was easier to use and learn for drawing shapes (despite some speech recognition errors). Sedivy and Johnson (1999) investigated a multimodal approach using speech input with a tablet and stylus to support sketching activities. The system utilised various speech commands for operations such as selecting tools, colouring, grouping, layering, scaling, and resizing pen, alongside the use of a stylus for drawing. Examples of speech commands supported include “thicker pen” and “thicker brush”, although the authors did not include examples of other speech commands supported by their multimodal drawing application. A user evaluation found that voice interaction supported participants’ creative process, shortened tool selection time, and reduced cognitive workload. However, some participants highlighted that background noise in their working environment reduced speech recognition accuracy, thus resulting in them having to repeat commands on multiple occasions to perform an action. Users also needed to memorise speech commands as the supported set of voice commands were not displayed within the interface. Alsuraihi and Rigas (2007) compared a multimodal speech approach (a combination of voice recognition and mouse) with a traditional mouse across a range of standard design activities (e.g., creating buttons, choosing colours, writing text, and the selection of different tools). The findings suggested that voice recognition reduced reliance on traditional inputs for selection of drawing tools, although speech recognition errors caused difficulty in accurately completing the tasks. The authors also did not explicitly provide the list of commands that were used for design activities. Van der Kamp and Sundstedt (2011) examined the use of voice input with eye gaze where voice commands (“start, “stop”, “snap”, “open colours” etc.) were used for selecting tools and initiate drawing and eye gaze for positioning the mouse cursor on a design canvas where user wanted to draw shapes. Overall, results highlighted that a gaze and speech combination supported a more efficient drawing process than using a mouse and keyboard, as well as presented a more engaging experience for participants. Qualitative feedback from participants highlighted that the speech commands were easy and straightforward to utilise for tools selection and drawing operations. Although participants also mentioned that they felt the use of a mouse was more accurate than eye gaze which caused inaccurate clicks on occasions when positioning the cursor on canvas.

Laput et al. (2013) presented the PIXELTONE application where direct manipulation (via touch) is used to select parts of an image, along with a limited set of high-level voice commands to perform contextual image editing operations (e.g., applying filters). Whilst a limited set of terms were provided for image transformation (“shadows”, “left”, “brighter”, etc.), results from a user evaluation found that the use of voice input presented interaction benefits over a “touch-only” version of the application. However, the authors also perceived that comparative speech commands such as ‘brighter’ were misleading and caused potential misinterpretation and confusion about the precision of brightness level. Similarly, Srinivasan et al. (2019) presented a multimodal approach using speech interaction and touch input for image editing operations which worked based on a recommended set of speech commands related to image editing actions displayed on a suggestions panel. Touch input was used to select interface elements and voice commands to perform image editing operations (e.g., “change fill color”, “add a sepia filter”). Results highlighted overall positive perceptions from participants, although there were also some usability challenges. For instance, the system could not appropriately display the recommended voice commands (e.g., for changing border size) due to the complex interface design (where other interface elements and tools obscured the suggested commands panel). Moreover, multiple speech commands recommended by the system were not perceived as intuitive by some participants and impeded their decision making in choosing an appropriate speech command to perform an action (hence contributing to increased cognitive workload). Kim et al. (2019) also utilised a multimodal approach using a stylus pen in conjunction with speech input. In particular, the authors investigated the use of short vocal commands in creative applications to support expert designers (e.g., “brush” to select the brush tool). A user evaluation found that these short voice commands helped creative experts access various design features more efficiently, thus helping to reduce cognitive and physical load. Previous research has also investigated the potential of non-verbal speech interaction to support people with physical impairments in producing freeform drawings – for instance, Harada et al. (2007a; 2007b; 2009) explored the use of a vocal joystick that enables continuous voice input in the form of vowel sounds to guide drawing directions (e.g., sounds like “aaa” for up and “ooo” for down). However, the authors highlighted interaction challenges with this approach in relation to smooth mapping of vowel sounds to the movements of a brush tool. Furthermore, it would be challenging to manipulate different shapes on a design canvas using continuous vowel sounds when variable transformation speed and sizes are required to precisely position, resize, or rotate objects.

As highlighted in the previous work covered, the majority of studies in this area have tended to investigate multimodal approaches where speech has been utilised in combination with other interaction methods. However, there are key challenges associated with this type of approach such as having to configure multiple devices, as well as the additional cognitive load due to operating additional input tools (Song et al., 2012). The use of additional input modalities such as mouse, keyboard, and gestures (Alsuraihi and Rigas, 2007; Alibay et al., 2017) also present significant challenges for people with physical impairments and can impede access to producing creative outputs (Creed et al., 2014). A unimodal interaction approach utilising speech only input can present an opportunity to address some of these limitations through providing a single point of input that can potentially reduce interaction complexity and cognitive load (Reitmaier et al., 2022). It also supports hands-free interactions through reducing the reliance on physical assistive tools such as ergonomic mouse and keyboard, and mechanical switches (Cave and Bloch, 2021). Whilst previous studies have explored the potential to utilise voice control within creative domains to support a range of creative design activities, a fundamental area where there has been limited research to date is around how digital assets can be efficiently manipulated (i.e., positioned, resized, and rotated) using this method of interaction.

Object manipulation within a creative design environment is a fundamental aspect of designing interfaces and typically involves core activities such as moving, resizing, and rotating digital objects on a design canvas. The object manipulation requires the precise control over creative assets to enhance the aesthetic appearance of visual designs (Xu et al., 2014). There is currently a gap in the literature around more nuanced positioning approaches that can support both rapid and accurate movement of objects across a large design canvas, as well as methods that easily facilitate pixel-level manipulations. The previous studies highlighted have tended to examine only basic approaches for positioning of objects through only directional commands such as “left”, “right”, “up” and “down”. Similarly, in terms of resizing of objects, relatively simple approaches have been explored to date such as the use of commands such as “bigger” and “smaller” to transform an object’s properties. Further research is required to explore alternative methods that provide users with physical impairments more control over adjusting an object’s size, especially in terms of how common approaches used in mainstream applications can potentially be tailored for voice control (e.g., the use of alignment guides and snapping features – (Baudisch et al., 2005; Dellisanti et al., 2008; Ciolfi Felice et al., 2016; Heo et al., 2012; Masui, 2001)). Finally, no work in the literature to date has examined methods for efficiently rotating objects via voice control on a design canvas. This is another fundamental action that designers need to regularly perform, although it remains unclear which commands are most suitable to support this activity and which interaction challenges may need to be overcome to facilitate the effective rotation of objects. Similarly, previous research studies have typically focused on the use of simple transformation commands (e.g., “left” and “right”), although these approaches do not typically support precise control of digital assets (e.g., the ability to move an object to a desired location via a specific number of user-defined pixels) (Gygli and Ferrari, 2020). Furthermore, there is a lack of work around exploring speech interaction approaches with the involvement of participants who have a diverse range of physical impairments, thus making it challenging to determine whether results highlighted in the literature can be generalised for physically impaired users. It is therefore crucial that further research now focuses on thoroughly exploring different speech only interaction techniques that facilitate fundamental visual design activities (i.e., object positioning, resizing, and rotation) to ensure optimal approaches are developed that support creative design workflows for people with physical impairments.

2.5 Summary

This chapter initially presented and discussed the nature of disability and different forms of upper body physical impairment, as well as the types of assistive tools and technologies that are used to support control of digital systems. A literature review was also presented that focused on highlighting research into different speech interaction techniques that can support people with physical impairments, as well as work focusing specifically on speech interaction to support creative visual design. In particular, a significant gap was identified around a lack of work exploring the use of speech input as a primary interaction method for supporting fundamental object manipulation activities. To address this gap in the literature, the following three chapters present a series of studies focused around novel voice controlled interaction approaches that facilitate object positioning, resizing, and rotation of digital assets.

3 STUDY 1: OBJECT POSITIONING USING SPEECH INTERACTION

This work has been published in the proceedings of 23rd ACM conference on International Conference on Multimodal Interaction as: "Aziz, F., Creed, C., Frutos-Pascual, M. and Williams, I. (2021) Inclusive Voice Interaction Techniques for Creative Object Positioning. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 461-469".

3.1 Introduction

The movement and positioning of digital objects around a design canvas is a fundamental object manipulation activity associated with creative visual design work (Kwan and Betke, 2011). By creative visual design, this work refers to user interface design which involves manipulating interface elements such as images, objects and shapes, typography, and interactive prototyping within a two-dimensional environment. Object movement and positioning across mainstream creative visual design applications (i.e., Photoshop Illustrator, XD, and Figma) is typically achieved through dragging objects via a mouse, manually adjusting the x and y values of an object via a properties panel, or by using navigational keys on a keyboard. However, as highlighted in Chapter 1, this type of method presents significant interaction challenges for people with physical impairments and can lead to them being excluded from using industry standard creative applications such as Photoshop, Illustrator, Figma and XD. Voice control holds significant potential to make visual design work more accessible (Hu et al., 2011), although there has been a lack of work examining how to use this interaction paradigm to efficiently position objects around a design canvas. Previous research has explored multimodal approaches that utilise speech to support object positioning (e.g., Hauptmann, 1989; Hiyoshi and Shimazu, 1994; Elepfandt and Grund, 2012), although these studies only presented initial investigations and did not conduct formal testing including people with physical impairments. As such, there has been limited research to date that has investigated the use of speech as the primary method of interaction for object positioning in creative work. Hence, by exploring speech controlled object positioning, this study would provide deeper understanding around the use of different speech interaction techniques which could support people with physical impairments in positioning objects on a design canvas for producing creative design work. This chapter presents work conducted to address this gap in the literature by investigating different speech interaction techniques to support positioning of graphical objects (i.e., images) around a design canvas. Two user studies are reported which explored object positioning via speech interaction – the first study investigated whether non-disabled users were able to position objects efficiently and effectively using three speech interaction techniques (Speed Control, Location Guides, and Positional Guides). This study provided important insights into the speech interaction approaches developed and highlighted opportunities for iterative improvements (prior to evaluating with physically impaired participants). A second user study was then conducted with disabled participants to validate the optimal speech interaction approach for object positioning. This chapter therefore presents three primary contributions: (1) three novel speech interaction approaches for positioning graphical objects; (2) two user evaluations presenting new insights into the use of speech interaction for object manipulation; and (3) validation of the Location Guides approach in supporting people with physical impairments to efficiently control objects on a design canvas.

3.2 User Study 1

3.2.1 Research Question

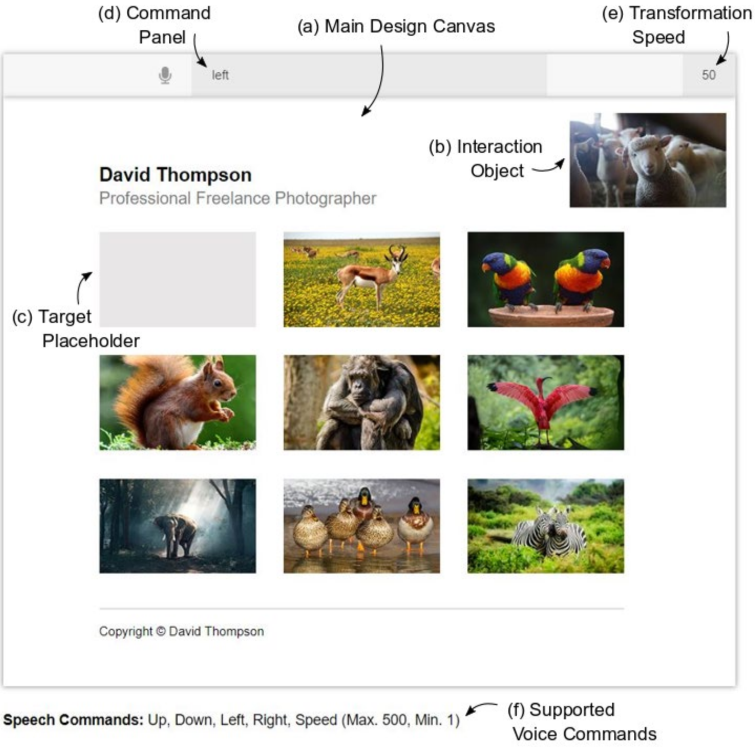
The following core research question was formulated to address the lack of work in this area:

RQ: How can voice interaction support the positioning of digital objects around a creative design canvas?

The open framing of this question intentionally encouraged research that was more exploratory in nature to help obtain a wider understanding of how new interaction techniques could potentially support the placement of creative assists with a design scenario.

3.2.2 Prototype Design

A new web-based research prototype was developed to investigate different object positioning approaches within a design canvas using speech interaction. The application was built using HTML, CSS, and JavaScript – with the Web Speech API (Web Speech API, 2023) used for detecting speech input from users.


left

50

David Thompson
Professional Freelance Photographer

Copyright © David Thompson

Speech Commands: Up, Down, Left, Right, Speed (Max. 500, Min. 1)

Figure 3.1 The main research prototype interface illustrating (a) the design canvas, (b) the interaction object, (c) the target placeholder, (d) speech commands panel, (e) transformation speed, and (f) a list of supported voice commands

The prototype simulates a 2D design canvas and displays a portfolio website design for a fictional professional photographer (Figure 3.1). The main design canvas (Figure 3.1 (a)) is fixed to 700x560 pixels to ensure that it can run on modern browsers without the need for vertical or horizontal scrolling (to help avoid inconsistent user experiences during evaluation studies). The design canvas consists of nine thumbnail sized images (150x90 pixels) which is similar in size and dimensions to those used in mainstream social media platforms to display posts and profile pictures (e.g., Facebook, LinkedIn, and Instagram). One of these images is the interaction object (Figure 3.1 (b)) that is presented in the top right corner of the design canvas (at the start of a task) and can be moved in all directions using a range of speech commands. The target placeholder (Figure 3.1 (c)) is displayed as a grey rectangular shape which represents the final position where an interaction object needs to be positioned. The placeholder is the same size as the interaction object (150x90 pixels) to ensure the image can be accurately placed over it. The speech command panel (Figure 3.1 (d)) is displayed at the top middle section of the screen to present and display spoken voice commands which the system has recognised. Switch input (e.g., a keyboard, mechanical switch, head tracker, foot pedal, etc.) can be utilised for initiating the speech recogniser on each task across the three interaction approaches. Audio feedback (a popping sound effect) is also played after a voice command has been issued to make the user aware that their input has been recognised.

The transformation speed (Figure 3.1 (e)) can be seen in the top right corner and is used to facilitate users in moving interaction objects slower or faster depending on a user's preferences. The transformation speed can range from values of “1” to “500” and is controlled using the voice command “speed” followed by a number (e.g., “speed 10”). For instance, when a user wants to move the interaction object towards the left and the transformation speed is set to 10, the image will move 10 pixels to the left when the appropriate voice command is issued (e.g., “left”). The transformation speed can also be used to fine-tune the interaction object position when it is closer to the target placeholder (selecting the minimum speed value, e.g., “speed 1”). The user can move the interaction object based on pixel values as opposed to continuous animation at the set transformation speed - this decision was taken as latency in processing of speech recognition can result in slight delays of commands being issued, which in turn can lead to objects moving beyond the user's intended target position (if continuous animation is used) (Karimullah and Sears, 2002). The supported speech commands (Figure 3.1 (f)) at the bottom of the design canvas are always visible to help users in recalling the available commands – these commands are dynamically changed on different screens depending on the object positioning approach used. The prototype particularly focused on three novel speech-controlled object positioning approaches to explore their potential in supporting object movement and positioning, as well as the challenges associated with these approaches. The following subsections present the three different object positioning approaches developed – Speed Control, Speed Control + Location Guides, and Speed Control + Positional Guides, as well as rationale for choosing these speech-controlled positioning methods:

3.2.2.1 Speed Control

This approach uses simple voice commands to move objects (images) around the design canvas. These commands include “left”, “right”, “down”, “up”, and “speed *x*” (x is the transformation speed value in numbers) to move the interaction object presented in the top right position of the design canvas (Figure 3.1). These voice commands were informed and motivated from previous work in the field (e.g., Karimullah and Sears (2002) where directional voice commands were used to control different interface elements (e.g., a mouse cursor). This approach mainly relies on the transformation speed (as discussed above) to move and position the interaction object over the relevant placeholder. There are no additional visual cues or layout tools (in contrast to the other two approaches) to support image movement around the design canvas. Speed Control therefore provides a baseline to compare against the other two positioning approaches that utilise additional visualisations to support object positioning. Similarly, no previous research has explored the precise positioning of objects on a two-dimensional design canvas using short voice commands such as those presented for the Speed Control method.

3.2.2.2 Speed Control + Location Guides

This approach utilises the same speech commands used in the ‘Speed Control’ approach, along with the use of location guidance to assist image positioning around the design canvas. The location guides were presented as a grid of circular labels (numbered from 1-90) overlaid on top of design canvas. The location guides can be displayed using the command “locations” – objects can then be moved by stating the number contained within the label that is closest to target placeholder, thus resulting in the top-left corner of the interaction object being placed over the appropriate location guide (Figure 3.2). Location guides can also be hidden by stating the “hide” command – users can then simply use the Speed Control method and transformation speed to refine the specific location of an object.

Speech Commands: Locations, Hide, Up, Down, Left, Right, Speed (Max. 500, Min. 1)

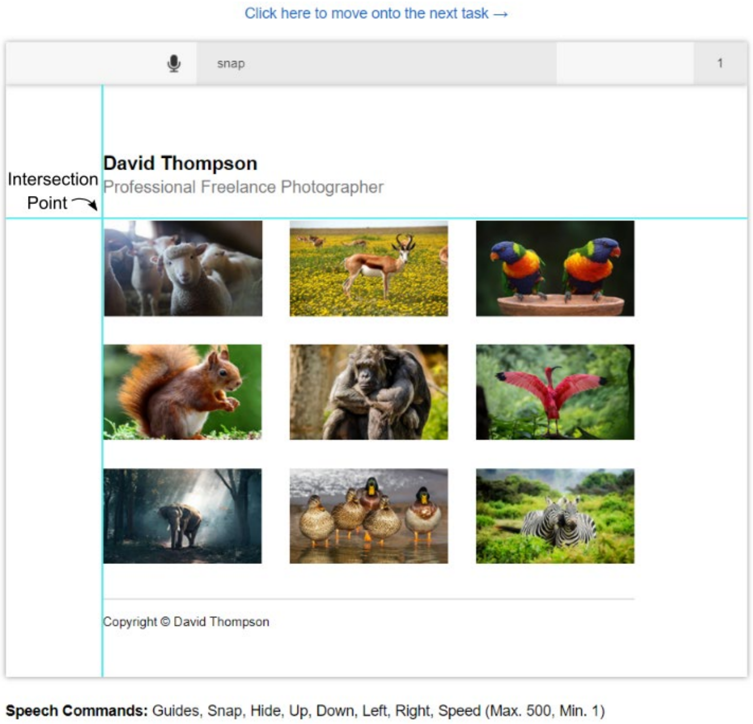
Figure 3.2 A screenshot of the 'Speed Control + Location Guides' approach where the top left corner of an interaction object (image) has been placed at the nearest possible position (Location 22) over the target placeholder

Informal tests with three field experts (i.e., professional interface designers) were conducted to evaluate a range of different label sizes and distances which helped to inform a balance between the size and distance of location guides to ensure users were not overloaded with options (which could result in a cluttered user experience). These participants were recruited through existing links within the university and were contacted via email to request their participation. This informal testing was conducted online using Microsoft Teams where participants initially opened a prototype link provided to them and then shared their screen content with the researcher. The testing tasks involved moving an interaction object over a target placeholder through utilising circular labels displayed in different sizes and distances. Label sizes that were evaluated included 15x15px, 25x25px, 30x30px, and 35x35px, while different distances between these labels were 65px, 70px, 80px, and 85px. During informal feedback, participants highlighted that the balance between different label sizes was not optimally supporting the aesthetic appearance and interaction within the overall interface design. In discussion with participants, it was suggested that label sizes of 20x20px placed 75px (≈2cm) apart would provide a better balance and interaction experience. This was tested and confirmed with participants during the informal testing and integrated into the prototype (Figure 3.2).

The opacity of the chosen interaction object is lowered when the location guides are displayed to ensure the grid of labels are not obscured by visual content on the canvas. This approach is motivated from Adobe XD (Adobe, 2023) where numeric labels are used for selecting tools and features via voice. The Location Guides approach is novel and different from the method used in Adobe XD as the circular labels are utilised for object positioning, as opposed to performing object and tool selection in the application. Furthermore, this type of approach has not yet been explored within the literature to support object positioning within 2D design environments.

3.2.2.3 Speed Control + Positional Guides

The third approach uses standard vertical and horizontal positional guidelines commonly used in mainstream design software (e.g., Adobe Photoshop and Illustrator). A similar approach has also been used in related work investigating the positioning of graphical objects within a design canvas via the combination of gaze interaction and mechanical switches to support people with physical impairments (Creed et al., 2020). However, Positional Guides presents a novel approach through enabling users to control the guidelines using voice-only interaction where objects can be positioned at the intersection of a vertical and horizontal guideline displayed on the canvas. In Positional Guides, users can initially enable the positional guide feature by stating “guides”. A single horizontal and vertical line are then displayed within the main canvas which can be moved through stating “left”, “right”, “up” or “down” (Figure 3.3). The transformation speed is also associated with moving guidelines and can be changed using the same speed commands as discussed previously. As a first step, the vertical and horizontal lines need to be moved so that their intersection point is at the top left corner of the user’s desired location. Use of the “snap” command then moves the interaction object to the point at which the lines intersect (Figure 3.3). The “hide” command can be used to hide the positional guides from the main view. Users can also utilise the Speed Control if the interaction object is not snapped at the exact position over the relevant placeholder. This approach is different than Location Guides in that it only uses the single vertical and a single horizontal guideline which can be moved on the canvas to adjust a single snap point at a time. Whilst Location Guides approach uses multiple circular labels and user can position the object to each visible circular point without moving and adjusting those labels.



Speech Commands: Guides, Snap, Hide, Up, Down, Left, Right, Speed (Max. 500, Min. 1)

Figure 3.3 An example of the Positional Guides approach where an image (top left) is snapped at the intersection of the lines using the 'snap' command

In creative design applications, both object selection and manipulation are required to move an object across a design canvas. It is important to note that the three speech interaction approaches perform object positioning in isolation and do not include the selection of creative assets. However, the selection and positioning of objects are two different types of manipulations – previous studies have widely explored object selection via speech interaction (Dai et al., 2005; Karimullah and Sears, 2002; Zhu et al., 2010), although there has been a lack of work investigating voice-controlled object positioning to date. The three approaches were therefore developed to work with pre-selected interaction objects to support a deeper exploration around perceptions of the positioning approaches as an individual manipulation (to ensure selection approaches did not bias participants' experiences).

3.2.3 Methodology

An online user evaluation was conducted with non-disabled participants to investigate whether they were able to position objects effectively via the three approaches developed. A within-participant design was utilised to ensure the approaches were viable and usable prior to running evaluations with participants who have physical impairments. Since the main criteria for controlling the research prototype is the use of voice input (regardless of whether a person has physical impairment or is able-bodied), it was believed that working with non-disabled participants initially would provide insights into the different approaches and opportunities for iterative improvements prior to working with physically impaired participants. It was also felt that there would likely be consistency between views of non-disabled users and participants with physical impairments (whose voice is not impaired) given the primary method of interaction (i.e., voice input) is consistent across both user groups. However, this is always important to validate as there may be a range of factors that could influence disabled users' experiences in using a voice-controlled interface. For instance, perceptions could be impacted by factors such as the potential need to integrate a voice-controlled system with other assistive technologies, physical impairments presenting challenges in initiating and controlling a speech recogniser, additional customisation of a voice-controlled interface to meet an individual's specific requirements, and personal views on the general efficacy of speech as a unimodal interaction modality. A follow-up study with people who have physical impairments is therefore detailed in Section 3.3).

3.2.3.1 Participants

Thirty non-disabled participants (15 male and 15 female) were recruited from a population of University staff and students via email. Participants were aged between 19 to 52 years (M= 28.13, SD=7.57) and were native English speakers. Participants self-identified their level of experience with graphical design software (10 Novice [no experience], 17 Intermediate, and 3 Experts), prototyping applications (15 Novice, 13 Intermediate, 2 Experts), and speech interaction technology (8 Novice, 16 Intermediate, and 6 Experts) (Appendix 8.1: Object Positioning [page 164]).

3.2.3.2 Apparatus

Since the experiments were conducted online, participants were required to use their own computer and microphone for voice input, as well as a keyboard for simulating a switch to control the speech recogniser (using the spacebar key). The Google Chrome browser (version 80 or above) was required for experimental tasks due to browser compatibility with the Web Speech API.

3.2.3.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.1: Object Positioning [page 160]). Initially, a recruitment email was circulated to university staff and students where those interested were required to contact the named researcher via email to schedule the testing session, as well as confirm their preferred platform for the evaluation session (e.g., Microsoft Teams or Zoom). At the start of testing sessions, the URL of the object positioning prototype was provided to participants which they were asked to access and then share their screen content. All necessary paperwork such as the participant information sheet and consent form were built into the prototype. Participants initially went through the participant information sheet (PIS) and had the opportunity to ask any questions to ensure they understood the purpose of the study and were willing to participate. Once this was confirmed, they were redirected to a formal consent form page where they read consent conditions and selected the available button to formally provide their consent to participate in the study. Participants were then presented with a pre-test survey asking questions around demographic information, graphic design and speech interaction experience (Appendix 8.1: Object Positioning [page 164]).

After completing the survey, participants moved onto the training task (moving a single object around a blank design canvas) to understand how to operate the relevant object positioning method before starting the main tasks for each interaction approach. Participants then clicked a link at the top of the screen to move onto the main evaluation which consisted of three interaction approaches tested via nine object positioning tasks (i.e., 27 tasks in total). The order of interaction modes and tasks were counterbalanced to minimise the potential for order bias. For each task, a different image was always placed in the top-right corner of the interface to be positioned over its associated placeholder (i.e., the grey box). Placing all interaction images in the top-right helped to ensure a consistent approach was adopted for comparing the different interaction techniques, as well as ensuring that participants had to move images in all directions to successfully complete tasks. Before starting a task participants activated the speech recogniser on each task screen using the spacebar key. After successfully positioning an interaction object over the relevant placeholder, participants selected the next task link located above the design canvas via mouse click.

Once all nine tasks had been completed for an interaction technique, participants were presented with the system usability scale (SUS) form (Bangor et al., 2009). They then started on the next technique with an initial training session, followed by the main tasks, and then the SUS form again. Participants were also asked semi structured interview questions at the end of testing session to explore their perceptions of the different interaction techniques (Appendix 8.1: Object Positioning [page 166]). Testing sessions lasted between 26 to 60 minutes in total.

3.2.3.4 Measures

Task completion time, positional accuracy, speech recognition performance, and SUS scores were calculated to evaluate the three interaction techniques. Task completion time was measured from when participants selected the start task link until they selected the next task link upon task completion. Positional accuracies (distances) were measured through task-wise arrangement of final interaction objects and 8 target placeholder locations, calculating the differences between these values, and then finding the Euclidean distance values from x and y values for each interaction approach (Wang et al., 2005). This was calculated using the following formula:

$$d(x, y) = \sqrt{\sum (x_n - y_n)^2}$$

Speech recognition performance was measured via the total number of speech commands, as well as speech recognition errors to compare between speech interaction approaches. Speech recognition errors were categorized as: ‘Speech Misrecognition’ – where the recogniser incorrectly interpreted a voice command (e.g., the speech command “right” misrecognized as “write”); ‘System Error’ – where system did not perform an action due to latency issues with the Web Speech API; and ‘Unsupported commands’ – where users issued commands that did not match those that were available to support the system. System usability scale (SUS) was used to evaluate perceptions of usability for each interaction approach (Bangor et al., 2009). Post-study questions also explored participants’ perceptions around each interaction approach, their overall impressions of using multimodal speech interaction for object positioning, and suggestions for improvements (Appendix 8.1: Object Positioning [page 166]).

3.2.4 Results

The Shapiro-Wilk’s (Shapiro and Wilk, 1965) normality test ($p > 0.05$) found task completion data to be normally distributed while positional accuracy, speech performance, and SUS data were not normally distributed. A one-way repeated measure ANOVA (Girden, 1992) was utilised to analyse the differences between task completion times for each interaction approach. A non-parametric Friedman test of differences (Zimmerman and Zumbo, 1993) for repeated measures was used with Bonferroni correction to analyse positional accuracy, speech performance, and SUS scores. Wilcoxon signed rank (Woolson, 2008) was used for post-hoc tests to further analyse differences in positional accuracy, speech performance, and usability scores.

3.2.4.1 Task Completion Time

A statistically significant difference was observed between Speed Control (M=11.33, SD=2.28), Location Guides (Mean=8.22, SD=1.92), and Positional Guides (M=12.39, SD=2.97) in relation to task completion time ($F(2, 58) = 49.16$, $p < 0.001$, partial $\eta^2 = 0.629$). Post-hoc Least Significant Difference (LSD) (Williams and Abdi, 2010) also showed a significant difference between Location Guides and Speed Control (sig = 0.001, $p < 0.05$), and between Location Guides and Positional Guides (sig = 0.001, $p < 0.05$). Figure 3.4 shows a box-plot demonstrating that participants took significantly less time to complete all tasks using Location Guides in comparison to the other two techniques.

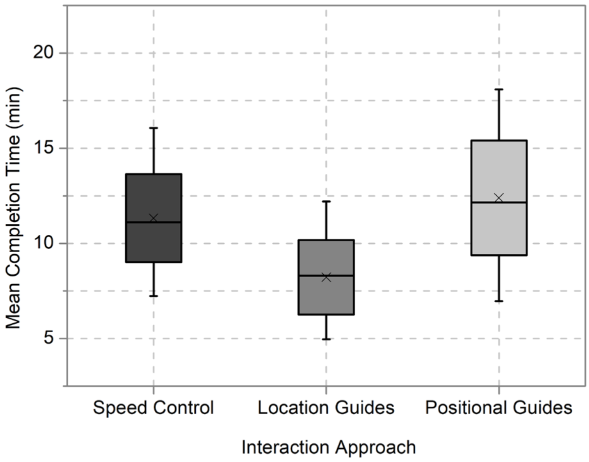


Figure 3.4. The mean task completion across three speech-controlled object positioning approaches

3.2.4.2 Positional Accuracy

The mean positional accuracy calculated via Euclidean distance across the Speed Control approach was 0.88 pixels (SD = 0.81), 0.62 pixels (SD = 0.49) for Location Guides, and 0.68 pixels (SD = 0.89) for Positional Guides. Non-parametric Friedman test results found significant differences in positional accuracy ($X2 = 0.014$, $df = 2$, $p < 0.05$). The post-hoc Wilcoxon signed rank highlighted a significant difference in positional accuracy between Location Guides and Speed Control ($Z = -3.92$, $p < 0.001$) and Positional Guides and Speed

Control (Z= -3.58, $p < 0.001$). However, no significant differences were observed between Location Guides and Positional Guides (Z = -0.69, $p = 0.48$). Figure 3.5 presents the mean positional accuracy calculated from Euclidean distances which demonstrates how close (accurately) the interaction objects were placed over their target placeholders. The lower value for Location Guides illustrates that participants were able to place interaction objects more accurately over target placeholders than Speech Control and Location Guides.

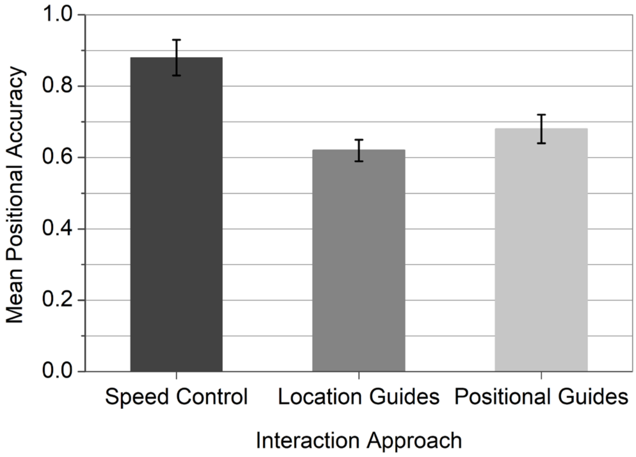


Figure 3.5. Mean positional accuracy across object positioning approaches

3.2.4.3 Speech Performance

The total number of speech commands issued across all 30 participants for Speed Control were 4596 (SD = 39.46), 3941 (SD = 31.96) for Location Guides, and 4837 (SD = 20.28) for Positional Guides. Non-parametric Friedman test results found statistically significant differences in the total number of speech commands issued ($X^2 = 14.63$, $df = 2$, $p < 0.05$). Post-hoc Wilcoxon signed rank highlighted statistically significant differences between Speed Control and Location Guides (Z = -2.39, $p < 0.05$), and Location Guides and Positional Guides (Z= -3.98, $p < 0.05$). However, no significant differences were found between Speed Control and Positional Guides (Z= -1.61, $p = 0.10$). The graph in the Figure 3.6 demonstrates that overall participants used fewer speech commands with Location Guides as compared to Speed Control and Positional Guides.

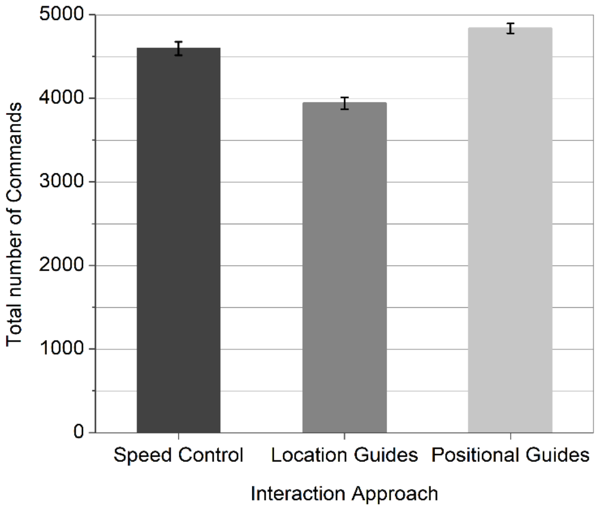


Figure 3.6 Total number of speech commands

In terms of speech recognition errors, there were a total of 281 (6.11%) speech commands related to 'Speech Misrecognition' for Speed Control, 240 (6.08%) for Location Guides, and 297 (6.14%) for Positional Guides. Friedman test results showed no statistically significant differences for 'Speech Misrecognition' across all three approaches ($X^2 = 0.10$, $df = 2$, and $p > 0.05$). There were 102 (2.21%) commands issued for Speed Control that were classified as 'System Errors', 80 (2.02%) for Location Guides, and 112 (2.32%) for Positional Guides. Friedman test results found no significant differences across the three interaction approaches ($X^2 = 0.12$, $df = 2$, and $p > 0.05$) in relation to 'System Errors'. There were 20 (0.43%) commands issued for Speed Control that were classified as 'Unsupported Commands', 15 (0.38%) for Location Guides, and 25 (0.49%) for Positional Guides. Friedman test results again showed no statistically significant differences across the three approaches ($X^2 = 0.77$, $df = 2$, and $p > 0.05$) in terms of 'Unsupported Commands'.

3.2.4.4 Usability Evaluation

The mean SUS score for Location Guides was 86.56 (SD = 14.09) and is rated as "Excellent" according to Bangor et al. (2009). Speed Control (M=76.83, SD=14.99) and Positional Guides (M=80.25, SD=14.27) can also be labelled as exhibiting "Good" usability (Bangor et al., 2009). Significant differences were found across the three interaction approaches ($X2 = 0.001$, $df = 2$, $p < 0.05$). The post-hoc Wilcoxon signed rank found a significant difference between Location Guides and Speed Control ($Z = -3.03$, $p < 0.001$). A significant difference was also observed between Location Guides and Positional Guides ($Z = -2.42$, $p < 0.001$). No significant differences were found between Positional Guides and Speed Control (Z = -

1.33, p = 0.19). The mean SUS scores across the three interaction approaches can be seen in Figure 3.7 demonstrating that Location Guides was rated as more usable than the other two interaction approaches.

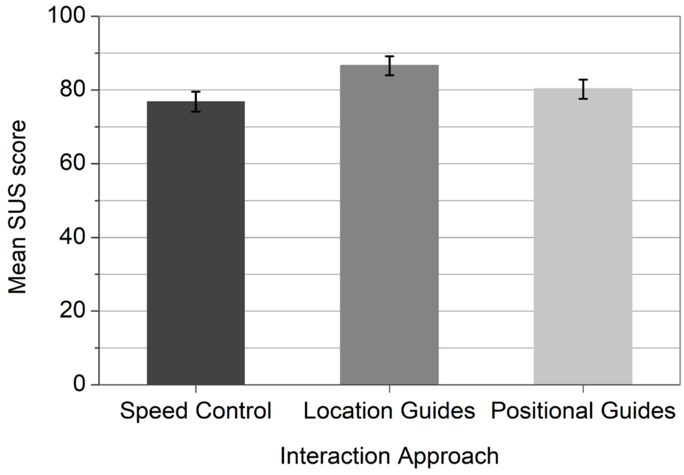


Figure 3.7 Mean SUS score across three object positioning approaches

3.2.4.5 Qualitative Feedback

In terms of feedback received from the semi-structured interviews, all participants provided positive feedback in terms of using the three object positioning techniques. For instance, Participant 24 commented that they "*…liked all methods, straightforward to use, not much training required, pretty intuitive*". Twenty participants highlighted an overall preference for Location Guides, seven preferred Positional Guides, while three participants preferred Speed Control.

In terms of Location Guides, participants who preferred this approach recognised that it was the most efficient and easy to use over the other methods:

"…location guides is easy to understand, and quickest, as you are able to get close to the position you want then just fine-tune at that location. I also like visual representation of these labels because this made easy to get target position very quickly". (Participant 13)

Participants who had a preference for Positional Guides tended to highlight that it was more accurate and reliable:

"I liked how precise Positional Guides method enabled me to place objects. It was easier to position image by aligning it based on guidelines". (Participant 17)

However, four participants explicitly commented that this method was slower than other techniques (e.g., Participant 19 commented that positioning took longer due to moving the guidelines and then have to issue the "snap" command).

Participants who preferred Speed Control commented that it was simple, easy, and straightforward to use. However, some concerns were also raised:

“Speed Control method had simple and fewer commands, but it required frequently repeating these commands to get the image to the correct place because of estimation problem when using speed values”. (Participant 7)

Four participants commented that the system occasionally interpreted their utterances incorrectly – for instance, the “right” command was identified as “write” or “wright”, “left” as “lift”, and “snap” as “nap”. Participant 6 commented that the speech command panel (i.e., (Figure 3.1 (f)) currently displays everything that a user verbalises (including general conversational speech unrelated to controlling the application) and suggested filtering this to only display recognised commands. No participants highlighted speech recognition or the use of a keyboard to control the speech recogniser as a significant issue in completing the evaluation tasks despite having a range of technical and graphical manipulation experiences.

3.2.5 Discussion (User Study 1)

Location Guides was found to be significantly faster and more usable for the image positioning tasks than Positional Guides and Speed Control. Participants were also more accurate in positioning images when using Location Guides as compared to Speed Control (although there were no significant differences between Location and Positional Guides). Subjective feedback also highlighted that the majority of participants preferred Location Guides and provided positive feedback in terms of its efficiency and usability in placing images around the canvas. In particular, participants highlighted that it enabled them to initially place images relatively close to the target position and then nudge the object into the correct location. This, in turn, led to fewer voice commands being issued (as opposed to the other two approaches) and shorter task completion times. In comparison, participants took longer to complete tasks when using Positional Guides as they had to initially move and adjust guidelines via the available speech commands (prior to re-positioning an object). Speed Control provided no visual hints and guidance, resulting in users having to repeat speech commands on multiple occasions in order to correctly position objects at the required position. Participants also highlighted that they liked the visual presentation of Location Guides which they felt contributed towards a more efficient and effective experience than other two approaches. No participants highlighted issues related to the use of speech commands and in completing evaluation tasks having a range of technical experiences therefore, the findings suggest that the presented speech-controlled positioning approach can support users with a range of technical and graphic design experiences. The findings from this study demonstrate the viability of positioning digital assets via the speech-controlled approaches developed. This work builds upon previous object positioning research that has tended to focus on voice interaction as part of a multimodal approach (Hauptmann, 1989; Hiyoshi and Shimazu, 1994; Elepfandt and Grund, 2012). These research studies utilised basic commands such as “left” and “right” for manipulating the position of objects, whereas the approaches developed provide more control over manipulating the location of digital assets. However, it is important to note that this study was conducted with non-disabled participants, so it is key to also validate whether voice-controlled object positioning is appropriate for the primary target audience (i.e., people with physical impairments). This forms the focus of a follow-up study detailed in the following section.

3.3 Follow-up Study

Since Location Guides performed significantly better than other two approaches, it was therefore decided that the Location Guides approach would be taken forward for further development and evaluation work in partnership with participants having physical impairments. Iterative updates were made to the research prototype based on the feedback received from the first study – in particular, the speech recognition system was improved further through mapping the homophones detected in the first study with the relevant correct commands (e.g., “write” was mapped with “right”, etc.). Filtering was also applied on recognised speech input to ensure that only supported voice commands were displayed in the command panel. Since the Location Guides method had a high SUS score and was considered easy to use and efficient, no further major updates were made to the prototype.

3.3.1 Methodology

3.3.1.1 Participants

Six participants (3 male and 3 female) were recruited through online advertisements (via social media networks i.e., Facebook, Slack community, Twitter, and LinkedIn) and the Central London RSI Support Group (2023). Participants were aged between 27 to 49 years (M=38.33, SD=7.69) and all were native-English speakers. Participants self-assessed their level of experience with graphical design software, prototyping applications, and speech interaction technology (Appendix 8.1: Object Positioning [page 164]). Table 3.1 details participants' impairments, technical experience, and assistive tools used during the testing session.

3.3.1.2 Apparatus

The study was conducted online, so all participants used their own computer, a microphone for voice control, and were encouraged to use their chosen form of switch input (e.g., foot pedal, mechanical switch, head tracker, keyboard, etc.) to control the speech recognizer. The Google Chrome browser (version 80 or above) was required due to its compatibility with the Web Speech API. The user evaluations were also conducted online using the Zoom video conferencing platform as all participants stated a preference for using this software.

Table 3.1. Participants Details: RSI = Repetitive Strain Injury; MM = Muscular Myopathy; GD = Graphical Design (Software); IP = Interface Prototyping (Software); ST = Speech Technology; AT = Assistive Technology

ID	Age/ Gender	Physical Impairments	Condition Details	Technical Experience	Switch input used during testing session	SUS Score
P1	27 (M)	Tenosynovitis (Since 2020)	Wrist Pain; Joint swelling and stiffness; Difficulty in using fingers.	*GD*: Average; *IP*: N/A; *ST*: Dragon, Apple Siri; *AT*: Head Tracker, Foot pedal.	Speech + Foot pedal	92.50
P2	34 (M)	RSI (Since 2014)	Hand tremors; Shooting pain in hands and arms; Pain in wrists; Tingling sensation in fingers.	*GD*: N/A; *IP*: N/A; *ST*: Dragon software, Talon, Apple Siri, and Google Assistant; *AT*: Eye tracking, Vertical mouse.	Speech + Keyboard	97.50
P3	49 (F)	RSI (Since 2012)	Severe pain and discomfort in hands; Tiredness in shoulders and upper arms.	*GD*: N/A; *IP*: N/A; *ST*: Dragon software; *AT*: N/A.	Speech + Keyboard	82.50
P4	41 (F)	RSI (Since 2010)	Fatigue; Shoulder pain; Sore wrists occasionally; Pulsing pain in fingers.	*GD*: Expert; *IP*: Expert; *ST*: Dragon software; *AT*: Vertical mouse, Mechanical Switch.	Speech + Jellybean Switch	100
P5	33 (M)	MM (Since 2009)	Muscles weakness; fatigue; Lack of balance; Difficulty with walking without sticks;	*GD*: Average; *IP*: Average; *ST*: Apple Siri; Google Assistant; *AT*: N/A.	Speech + Keyboard	67.50
P6	46 (F)	RSI (Since 2000)	Wrist pain, Pain in shoulders and upper arms; tiredness.	*GD*: N/A; *IP*: N/A; *ST*: Dragon software; *AT*: N/A.	Speech + Keyboard	72.50

3.3.1.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.1: Object Positioning [page 160]). A recruitment advertisement was disseminated across social media networks (via Facebook, Slack, Twitter, and LinkedIn) where those interested were directed to contact the named researcher via email. Testing sessions were then scheduled for participants (by requesting their preferred date and time), as well as confirming which video conferencing platform they would prefer to use for the evaluation. During testing sessions, participants initially went through the PIS where the researcher ensured they understood the purpose of the study and that they were willing to participate. Once this was confirmed, they were redirected to a formal consent form page where they provided formal consent by selecting the available confirmation button. Participants were then presented with a pre-test survey including questions focused around demographic information, graphic design and speech interaction experience (Appendix 8.1: Object Positioning [page 164]). Participants then moved onto a training task in which they were able to re-position a single interaction object (image) across a blank canvas using the updated Location Guides approach. Once they were familiar with the positioning technique, they were asked to complete the same nine image positioning tasks (using only the Location Guides method). Participants then completed the SUS questionnaire, followed by some open-ended questions focused around their perceptions of the interaction approach in the context of object positioning (Appendix 8.1: [pages 165-166]). Evaluation sessions lasted between 20-35 minutes.

3.3.2 Results

Participants utilised a range of different switches to successfully control the speech recogniser with no major issues reported. Participant 1 has Tenosynovitis and utilised a foot pedal during the testing session (as a switch control) which they were able to effectively use for activating the speech recogniser. This participant was particularly complimentary about the efficacy of the tool for positioning of objects over larger distances and highlighted no major issues with the Location Guides approach. Four participants (Participants 2, 3, 4, and 6) self-disclosed a repetitive strain injury with three (Participants 2, 3, and 5) using the keyboard spacebar key as a switch, whilst the other (Participant 4) used a Jellybean switch during the testing session. All four of these participants had previous experience of using speech technology (i.e., Dragon software) and did not encounter any issues when enabling the speech recogniser using their chosen switch control. Participant 5 (who has muscular myopathy) also had previous experience with speech technology (i.e., using Apple Siri) and utilised the spacebar key as a switch control. Similarly, no major issues were reported by this participant in terms of controlling the speech recogniser using a range of assistive devices. It was clear from the results that despite participants having a range of technical experiences and designing experiences, no participants highlighted significant issues related to the use of speech commands and in completing evaluation tasks therefore there was consistency across the findings from both user evaluations.

A mean SUS score of 85.42 (SD=12.28) was received for the Location Guides method which can be labelled as ‘Excellent’ (Bangor et al., 2009) (Table 3.1). Three participants (Participant 1, Participant 2, and Participant 4) provided the highest scores (92 and above), whilst Participant 3 (82.50) and Participant 6 (72.50) scores were rated as ‘Good’. Participant 5 provided a lower score (67.50) which can be rated as ‘Average’ – who experienced some speech recognition issues on occasions during the study (i.e., the recogniser not detecting the participant’s voice input), although it is important to note that they were still able to successfully complete all tasks. It is important to note that the usability score of Location Guides across initial user evaluations and follow-up study were consistent as both groups provided scores that can be labelled as “Excellent” according to Bangor et al. (2009). This highlights that the Location Guides is usable across a diverse range of participants having a variety of technical and design experiences.

The overall task completion times ranged between 6.24 (Participant 2) – 13.95 minutes (Participant 3), whilst the mean task completion time across all six participants was 10.93 minutes (SD = 3.18). The positional accuracy was again calculated using Euclidean distances (Wang et al., 2005) for individual participants where mean values ranged from 0.54 – 1.22 pixels (M=0.84, SD=0.25). A total of 597 voice commands were issued across all six participants where 34 (5.69%) were related to ‘Speech Misrecognition’, 10 (1.67%) commands were related to ‘System Errors’, and only 1 (0.16%) command was classified as an ‘Unsupported command’.

All participants provided positive feedback and were able to utilise the features within the research prototype for image positioning. Four participants (Participants 1, 2, 3, 6) commented that they liked the grid presentation of the location guides and that they enabled them to efficiently position objects:

“… I like that it gives you the control to easily do large imprecise movements followed by doing small more precise adjustments”. (Participant 1)

Additional positive comments about the positioning approach included:

"… is straightforward and very helpful to align objects because when you are not using mouse moving things around canvas is a difficulty". (Participant 3)

"… it could be an interesting and useful activity for dragging things on the interface for manipulation work". (Participant 4)

Participants also provided positive responses around using transformation speed to adjust the positioning of an object. For instance, Participant 5 commented that "... *with a little practice I was able to assess how many pixels of speed I needed to move image correctly, and gradually I was sure if I need to adjust image 1 or 2 pixels*". Two participants were also able to effectively utilise their own assistive technologies to complete the tasks (i.e., Participant 1 used a foot pedal for controlling the speech recogniser, whilst Participant 4 used a Jelly Bean switch).

3.4 Conclusions

This chapter has addressed the research question posed in Section 3.2 through presenting and evaluating a new system that facilitates three different methods of positioning graphical objects via speech interaction (Speed Control, Location Guides, and Positional Guides). The majority of participants from the first study found all three approaches to be viable, although Location Guides was perceived to be more efficient, and usable. People with physical impairments also provided positive feedback with results demonstrating that they were able to successfully complete image positioning tasks during the follow-up study.

This work therefore contributes a deeper understanding around the viability of speech interaction to make core tasks associated with creative activities more accessible. In particular, it moves beyond the basic positioning approaches investigated in other studies (Elepfandt and Grund2012; Hiyoshi and Shimazu, 1994; Karimullah and Sears, 2002) and presents new opportunities for designers with physical impairments to manipulate the location of digital assets. This work also demonstrates how common features typically utilised within mainstream applications (e.g., guidelines) can be tailored for voice interaction to support object positioning. Moreover, the findings from this research may have wider applicability in other related domains (e.g., positioning of objects in commercial office software – word processors, presentation applications, operating systems, etc.), as well as the potential to support object positioning on different platforms (e.g., tablets and mobile phones). Wider limitations and implications of this study are presented Chapter 6 (Conclusions and Future Work) alongside the research conducted in upcoming chapters.

4 STUDY 2: OBJECT RESIZE USING SPEECH INTERACTION

This work has been published in ACM conference on Designing Interactive Systems Conference (DIS`22) as: "Aziz, F., Creed, C., Sarcar, S., Frutos-Pascual, M. and Williams, I. (2022) Voice Snapping: Inclusive Speech Interaction Techniques for Creative Object Manipulation. In *Designing Interactive Systems Conference* (pp. 1486-1496)."

4.1 Introduction

This chapter details a research investigation into the resizing of digital objects using voice input. Similar to object positioning, resizing of graphical assets is an essential and fundamental manipulation activity associated with digital visual design work. For instance, in mainstream design applications, users can commonly control the shape and dimensions of objects through dragging transformation handles via a mouse or entering numerical values of width and height using a properties panel (Raisamo and Raiha, 1996; Frisch et al., 2011; Xu et al., 2014). However, whilst this is a common tool that is widely used across different applications, it requires the use of dragging movements that do not clearly map to common speech interaction techniques. Similarly, object snapping is an important and relevant technique that is commonly used in creative design applications to support users with precise alignment and resizing of digital objects (Bier, 1986; Masui, 2001; Heo et al., 2012; Fernquist et al., 2011; Augstein et al., 2018). Snapping approaches (via mouse control) typically involve 'smart' (sticky) snapping in alignment with other objects on the canvas or to guidelines that have been manually placed by users on the design canvas (Baudisch et al., 2005; Ciolfi Felice et al., 2016). Research has shown the use of snapping can support alignment and unity within designs (Xu et al. 2014; Creed et al., 2020), thus supporting a designer's workflow and the production of professional outputs. However, it is also unclear whether speech as a primary method of interaction can be beneficial for object snapping to support people with physical impairments in resizing creative objects. Previous literature around object resizing have only supported basic speech commands to manipulate objects – for instance, Sedivy and Johnson (1999) used commands such as “thicker pen” to resize pen stroke, and Williams et al. (2020) utilised “shrink”, and “enlarge” commands to resize 3D objects. However, these studies have presented multimodal interactions and are also not formally evaluated with people who have physical impairments. This chapter addresses the limited work in this area by presenting three primary contributions: (1) the development of new speech interaction approaches for resizing graphical assets informed through well-established object manipulation techniques; (2) a user evaluation with people who have physical impairments presenting new insights around the use of speech interaction for object manipulation (i.e., object resizing); and (3) research findings evidencing that automated object snapping for resizing actions (in voice control scenarios) presents interaction benefits in terms of usability and efficiency.

4.2 User Study 1

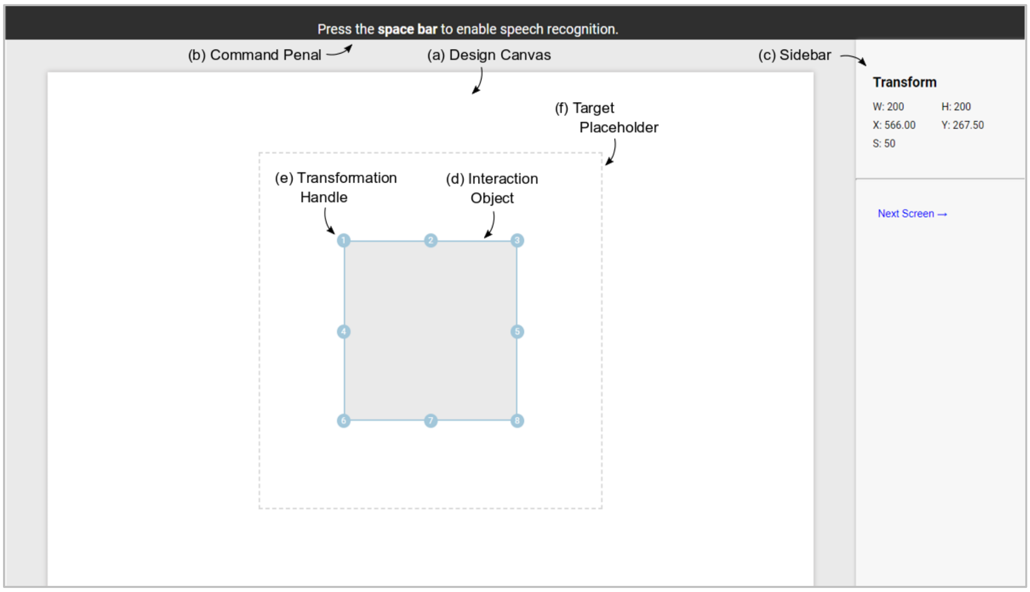
4.2.1 Research Question

The following overarching research question was formulated to address the limited work in this area:

RQ: How can voice interaction facilitate the efficient resizing of digital objects?
Similar to the previous chapter, the open nature of this question encouraged an exploratory approach to support better understanding around the viability of object resizing via voice control and the potential of new techniques to enable efficient transformation of digital assets.

4.2.2 Research Prototype

An object resize prototype was developed using HTML, CSS, and JavaScript (with the Web Speech API utilised for recognition of voice input). Figure 4.1 shows the research prototype interface which consists of a design canvas, speech command panel, a sidebar, an interaction object (grey box) with circular transformation handles, and a dotted target placeholder. The prototype dimensions were fixed to 900x600 pixels in size to reduce the potential for vertical and horizontal scrolling being required during user studies (which could lead to inconsistent user experiences). The spoken voice commands issued by users (Figure 4.1 (b)) are displayed within a black toolbar at the top of the prototype. Switch input (i.e., use of the spacebar key) was used for initiating the speech recogniser to enable recognition of voice commands. A popping sound effect was also played to provide audio feedback after a voice command has been issued to make the user aware that their input has been recognised. A sidebar (Figure 4.1 (c)) on the right-side is used to display the width, height, *xy* positions and transformation size values to keep users aware of the properties of objects they are manipulating. The interaction object (Figure 4.1 (d)) has eight transformation handles represented as circular labels with fixed numbers from 1 to 8 (starting at top-left to bottom-right). The interaction object can be resized through voice commands associated with a specific direction – for instance, "*x* big" or "*x* small" (where *x* is the number of the transformation handle from 1 to 8). The motivation for choosing "big" and "small" voice commands instead of "bigger", "smaller", "larger", "shorter", "increase" and "decrease" was to ensure commands are short, quick, and easy to pronounce (Kim et al., 2019). In addition, research has also suggested that shorter commands are more efficient and accurate in terms of speech recognition as compared to longer words and sentences (Elepfandt and Grund, 2012). Users can also manipulate the transformation size (in pixels) through the voice command "size *x*" (where *x* relates to the number of pixels – e.g., size 100). Target placeholders were also integrated into the design canvas and can be seen as a dotted box representing the target resize dimensions for the interaction object (Figure 4.1 (f)).



Press the **space bar** to enable speech recognition.

Transform

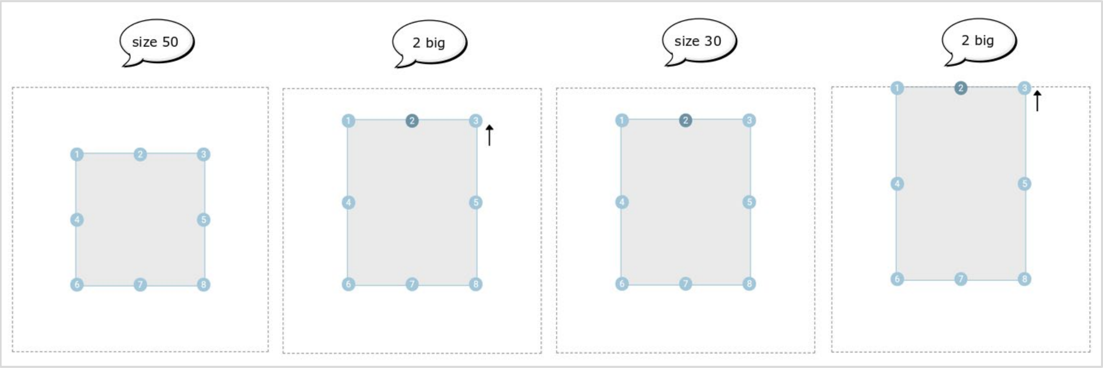
W: 200 H: 200

X: 566.00 Y: 267.50

S: 50

Figure 4.1 Object resize prototype (a) Design canvas, (b) Voice command panel, (c) Sidebar (d) Interaction object, (e) Transformation handle, and (f) Target placeholder

Eight object resize tasks were designed where the initial size of the interaction object was always kept consistent (200x200 pixels) while the dimensions of target placeholders was adjusted for each task. The size of target placeholders was based on covers, banners, profile pictures, icons, thumbnails, and logos available in social media applications such as Twitter, Facebook, Instagram, LinkedIn, Pinterest, and Tumblr (i.e., 800x450, 640x360, 400x400, 440x220 150x150, 77x77, 44x44, and 32x32). Figure 4.2 illustrates an example where an object is being resized from the top by changing transformation size values and issuing a "2 big" voice command.



(a) **(b)** **(c)** **(d)**

Figure 4.2. An example of a user attempting to resize the top of an interaction object to match with the target placeholder's dimensions - (a) User sets the transformation size through a "size 50" speech command, (b) user issues voice command "2 big" which resizes object from top, (c) the user then changes the transformation size via a "size 30" command, (d) the user again states "2 big" which resizes the object in alignment with the placeholder's top boundary

4.2.3 Methodology

A user evaluation was conducted online to gather insights around how participants found resizing objects on a design canvas via speech interaction. Similar to the study presented in Section 3.2, the key requirement for operating the prototype was the ability to use speech to issue commands (regardless of whether participants have physical impairments or not). It was therefore felt that initially evaluating with non-disabled participants would provide useful insights around the feasibility of the approach to ensure it would be viable prior to working with participants who had physical impairments.

4.2.3.1 Participants

The prototype was evaluated with 12 non-disabled participants (8 male, 4 female) who were recruited from a population of university staff and students via email. All participants were native English speakers aged between 20 to 48 years (M=30.33, SD=7.54). Participants also self-assessed themselves based on their level of experience with graphical design software (6 intermediate and 6 expert), prototyping applications (2 novice, 6 intermediate, 4 expert), and speech interaction technology (8 intermediate and 4 expert).

4.2.3.2 Apparatus

Studies were conducted remotely, so participants were required to use their own computer and microphone for voice input, as well as a keyboard for simulating a switch to control the speech recogniser (using the spacebar key). The Google Chrome browser (version 80 or above) was required for experiment tasks because of browser's compatibility with the Web Speech API.

4.2.3.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.2: Object Resize [page 167]). Testing sessions were scheduled with participants via email communication with sessions conducted online (via Zoom video conferencing platform). The researcher initially provided participants with a link to the research prototype which they were asked to access and then share their screen content. The prototype's first screen included the participant information sheet, followed by a consent form which they were required to read and then provide their consent through selecting a confirmation button. Participants then completed some pre-test questions requesting demographic information and experience with graphic design and speech technology (Appendix 8.2: Object Resize [page 170]). Once the survey had been completed, participants performed a training task where a single object was present on a blank white design canvas that could be resized using the voice commands available. After completing this initial task, participants then performed eight separate object resize tasks where they were required to resize an interaction object to the specific placeholder target size. Only one interaction object and target placeholder was displayed on each individual task screen – once a task had been completed, participants then moved onto the next task via a link presented in the sidebar (via a mouse click). After all tasks had been completed, participants were administered the SUS survey and asked to also complete an online survey with five open ended questions (Appendix 8.2: Object Resize [page 172]). The testing sessions were video recorded for later analysis with all testing sessions lasting between 20-30 minutes.

4.2.3.4 Measures

Task completion time, resize accuracy, speech recognition performance, SUS were used to evaluate the object resize approach. Task completion time was measured from when participants clicked the 'Next Screen' link in the right sidebar (which started the next task) until they then clicked the 'Next Screen' link again upon task completion. Resize accuracies (distances) were measured through calculating the Euclidean distance between the final width and height (in pixels) of both the interaction object and target placeholders (using the same formula highlighted in Section 3.2.3.4). To facilitate this calculation, the x and y positions of the interaction object and target placeholder were normalised, and the Euclidean distance between the top-right points of both objects was then calculated (Figure 4.3).

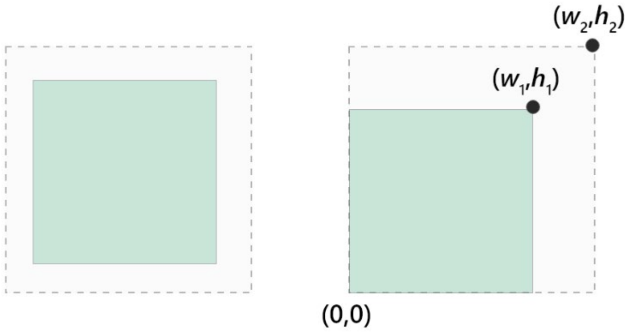


Figure 4.3 Euclidean Distance calculation to ascertain the differences in width and height of the interaction objects and target placeholders. The image on the left-side provides an example of the final dimensions of an interaction object (after completion of a task) when attempting to resize to the target placement dimensions (represented through the dashed border). The image on the right-side demonstrates how the x and y position of both objects has been normalised (i.e., to 0, 0 coordinates) to enable the Euclidean distance to be calculated between the top-right points of each object.

Speech recognition performance was again measured via the same three categories highlighted in Section 3.2.3.4: 'Speech Misrecognition', 'System Errors' and 'Unsupported commands'. SUS was used to evaluate perceptions of usability, whilst post-study questions explored participants' perceptions about the object resize approach, as well as suggestions for improvements (Appendix 8.2: Object Resize [pages 171-172]).

4.2.4 Results

The mean task completion time across all twelve participants was 11 minutes and 53 seconds (SD= 2.48), whilst the mean resize accuracy (calculated from Euclidean distances) was 0.93 pixels (SD = 0.96). A total of 1911 (SD =15.80) voice commands were issued across all participants where 236 commands (12.35%) were related to 'Speech Misrecognition', 34 commands were associated with ‘System Errors’, and 8 (0.42%) commands were classified as ‘Unsupported Commands’. The mean SUS score was 73.12 (SD = 9.42) across all participants which can be labelled as a “Good” level of usability (Bangor et al., 2009).

All participants provided positive feedback on the object resize approach stating that it was easy to use, simple and intuitive. However, three participants (Participant 4, 9, 11) highlighted that they felt frustrated when having to repeat the transformation handle label along with a “big” or “small” command (e.g., saying “1 big”, “1 big”, “...”, or “1 small, “1 small”, “…” to complete the same action). Six participants also highlighted that on occasions they had difficulty in estimating the correct transformation size:

“It takes some time to decide what size value should be selected. I wasn’t sure if I should select size 10 or size 12”. (Participant 5)

“… for smaller resizing I was still able to guess eventually how far 1 or 2 pixels will resize object but for bigger objects where longer jumps were required it was harder to guess correct size”. (Participant 6)

However, another participant highlighted that estimation became simpler after extended use:

“I was able to guess size value once I completed resizing one side then it was pretty easy to guess for the other side of object that how much pixels are required for correct resizing”. (Participant 2)

Four participants (Participants 3, 5, 11, 12) suggested that there should be some form of visualisation size such as a grid or a ruler which can help identify the correct transformation size required. It was also observed that some voice commands were not identified correctly – for instance, “big” was identified as “dig”, “beg”, “bag”, “wig” and “small” as “mol”, or “mall”, while “2” was identified as “too” and “4” was identified as “for”.

4.2.5 Study Discussion

Overall, results from this initial exploratory study found that all participants were able to successfully complete the eight resize tasks using speech input. In particular, resize accuracy results demonstrated that participants were able to accurately resize interaction objects in accordance with target placeholder dimensions (given that the mean Euclidean distance across all tasks and participants was less than 1 pixel). The voice-controlled approach was also perceived by participants has having a good level of usability, suggesting that the resizing approach may be a viable solution for people with physical impairments.

However, estimation of transformation size was a key issue reported by participants and is an area where further work is required. For example, Participant 6 commented that it was difficult to estimate the correct transformation size value where tasks required larger resize transformations, as compared to smaller ones. Moreover, participants also highlighted that visual hints could support them in correctly estimating appropriate transformation sizes. Based on both the quantitative and qualitative data collected, it was concluded that voice control for object resizing presented a viable approach and further developed our understanding around the potential of speech-controlled approaches. In particular, previous work around object resizing has typically focused on utilising multimodal approaches that incorporate voice input (Sedivy and Johnson, 1999; Kim et al., 2019; Williams et al., 2020), although no previous work to date has explored or validated the use of unimodal voice-controlled approaches to facilitate resizing of digital objects. This work therefore presents new insights and knowledge in this area, although further work is required to validate the findings with people who have physical impairments, as well as exploring whether alternative resize transformations could provide a more optimised experience.

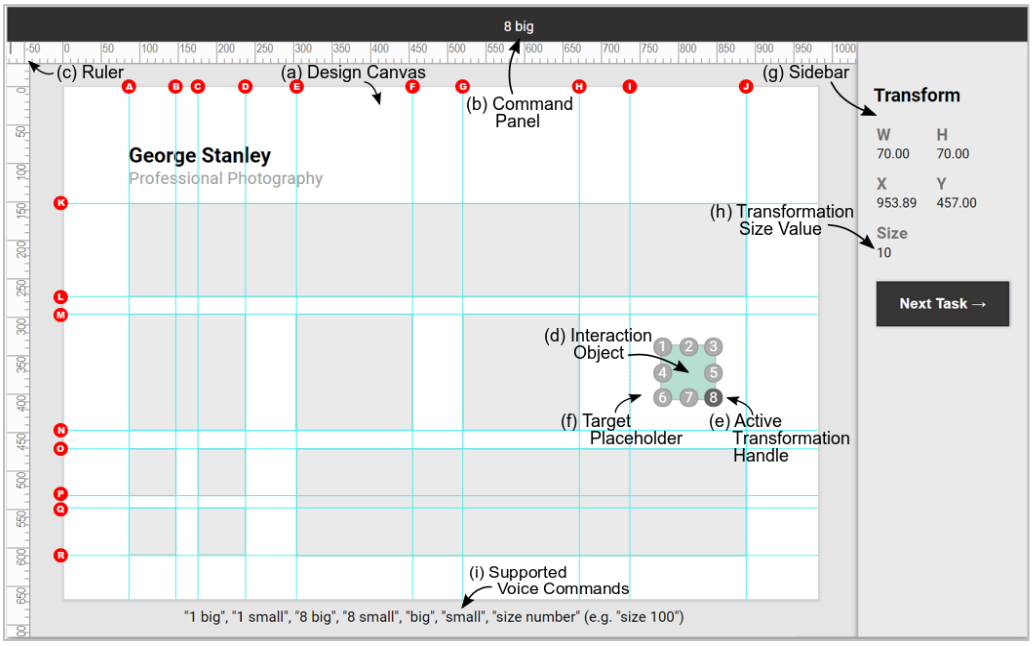
4.3 User Study 2

Whilst the exploratory study provided some important insights around the feasibility of resizing digital objects via voice control, it is also important to validate these findings with the core target audience (i.e., people with physical impairments). Iterative updates were therefore made to the prototype, in addition to the development of three different speech-controlled resizing techniques to investigate further the challenges identified around estimation of transformation size.

4.3.1 Research Prototype

In terms of iterative updates made to the previous prototype (utilised in the exploratory study), it was important to consider the issue highlighted by participants around having to repeat handle numbers with every “big” or “small” vocal command (which participants found tedious). To address this point, an update was made whereby users only need to state the handle they would like to manipulate once, followed by use of “big” or “small” commands (e.g., “1 big” followed by “big”, “big”, “small”, etc.). Homophones were also mapped with the relevant supported voice commands (e.g., “dig” or “wig” were aligned with “big”, etc.) to improve speech recognition performance.

The object resize prototype was again developed using HTML, CSS, and JavaScript (including the Web Speech API for speech recognition) and presented a typical creative design interface (Figure 4.4). The design canvas (Figure 4.4 (a)) contained a wireframe portfolio design mockup for a fictional professional designer comprised of common interface visual assets such as text and image placeholders of different sizes.



8 big

George Stanley
Professional Photography

Transform

W 70.00
H 70.00
X 953.89
Y 457.00
Size 10

"1 big", "1 small", "8 big", "8 small", "big", "small", "size number" (e.g. "size 100")

Figure 4.4. Research prototype interface – all guidelines (i.e., snap points) are displayed for reference, but are hidden by default

The speech command panel (Figure 4.4 (b)) was again displayed at the top of the screen as a black toolbar where recognised voice commands were displayed. A static ruler (Figure 4.4 (c)) was positioned at the top and left of the canvas with pixel values displayed on major ticks at 50 pixel intervals. This was integrated based on feedback from participants in the exploratory study where it was emphasised that a ruler could further support with challenges around estimating the size of object transformations. An 'interaction object' (Figure 4.4 (d)) was again displayed as a green shape with eight transformation handles around its borders represented as circular labels with fixed numbers from 1 to 8. The position of the interaction object was fixed and could not be altered while only the dimensions (width/height) of objects could be manipulated. Target placeholders for a specific task (visualised using a white background with a dotted border) were displayed in relation to the interaction object (e.g., Figure 4.4 (f)) and represent the final dimensions to which the object needs to be resized. The sidebar (Figure 4.4 (g)) contained the same common object attributes, whilst the supported speech commands (Figure 4.4 (i)) were again displayed below the canvas.

Three different object resizing approaches optimised for speech interaction were developed: ‘NoSnap’ (utilising only transformation handles) and two object snapping techniques (‘UserSnap’ and ‘AutoSnap’) which were focused around a common snapping feature in mainstream applications (i.e., Adobe XD and Figma). Figure 4.5 highlights this type of snapping approach where objects are resized through accessing a transformation handle via a mouse (or touch) and then dragging the object to the desired size. Whilst dragging, smart guides become visible which provide subtle visual hints for snapping the object in reference to other assets present on the canvas. Further details about each object resizing technique developed are provided in the sections below.

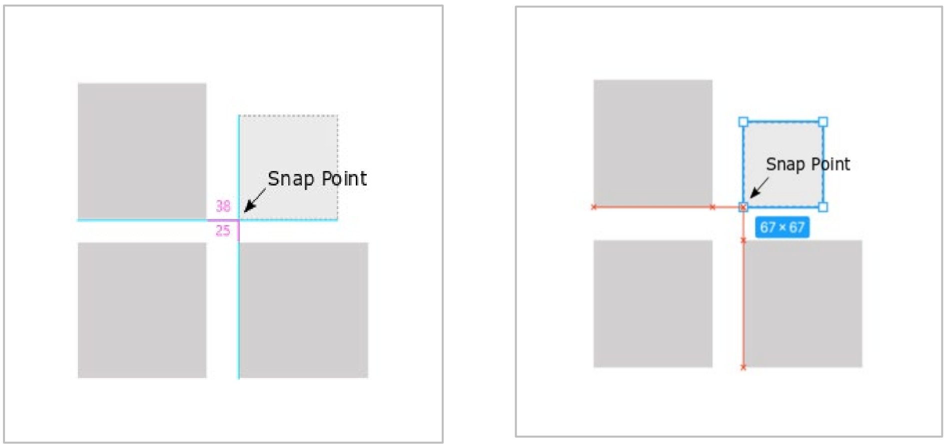
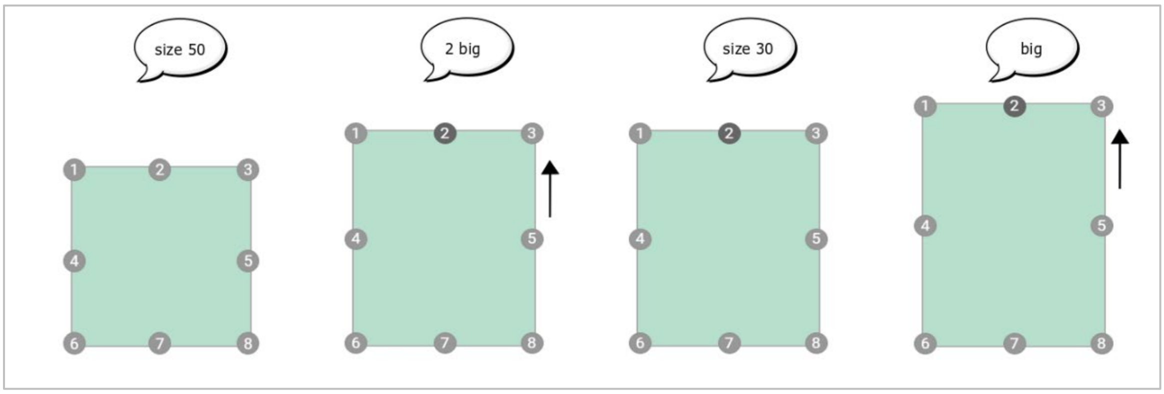


Figure 4.5 Smart snapping with Adobe XD (left) and Figma (right) – when the top-right object is resized using mouse dragging and it reaches the boundaries of the other three objects placed nearby, smart (sticky) snap guidelines become visible

4.3.1.1 NoSnap

This approach operates in the same way as the approach used in the exploratory study where users issue voice commands to specify and manipulate an object handle (e.g., “1 big”, “big”, “8 small”, “small”, etc.), as well as setting the transformation size by stating “size [number of pixels]” (e.g., “size 10”). For completeness, Figure 4.6 illustrates an example of the NoSnap approach where the top-side of a shape is extended – a “size 50” command is initially issued, followed by “2 big” to increase the object height by 50 pixels. The transformation size is then altered using a “size 30” command, followed by “big” to increase the object height by a further 30 pixels. A user can repeatedly issue the “big” or “small” command to continue manipulating a previously selected transformation handle. If a different selection handle is selected (e.g., “5 big”), the previous handle is deactivated and the new handle can then be adjusted.

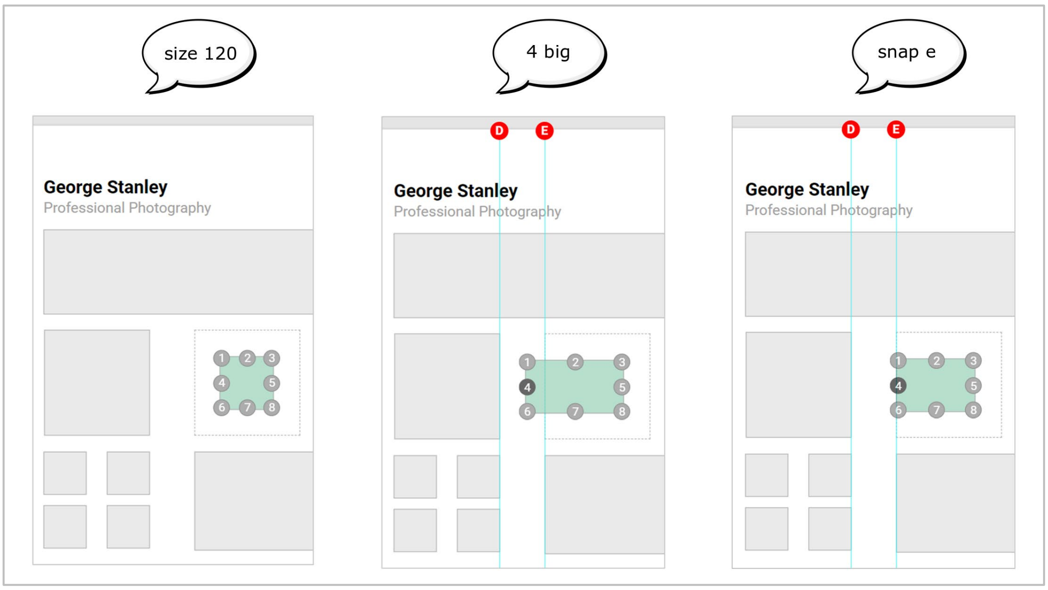


(a) **(b)** **(c)** **(d)**

Figure 4.6 (a) Transformation size set to 50 using speech command "size 50", (b) "2 big" command resizes the object from the top direction, (c) transformation size is again adjusted to 30 pixels using command "size 30", (d) "big" command is used to resize the object height from top direction

4.3.1.2 UserSnap

This approach combines the NoSnap features with object snapping in relation to nearby reference objects located on a digital canvas. A user can still resize an interaction object from any direction using the given voice commands ("1 big", "big", "8 small", "small" etc.), although a snap guide is displayed once the side of the object being manipulated is within a 100px threshold of a potential snap point. Each vertical and horizontal snap guideline is given a unique alphabetical identifier (A, B, C, etc.) displayed as a red circular label at the top and left edges of these guidelines. The user can then snap the object to the vertical or a horizontal guideline displayed using the voice command "snap x" (where x refers to the unique guideline identifier). The mock-up wireframe design consists of 10 vertical and 8 horizontal snap guidelines (Figure 4.4) – these are hidden by default and only guidelines within the 100 pixels threshold of the currently selected transformation handle are displayed. The threshold value was informed through previous research investigating mouse cursor snapping thresholds to support efficient target acquisition (Trewin et al., 2006). Figure 4.7 demonstrates how an interaction object can be resized via UserSnap.

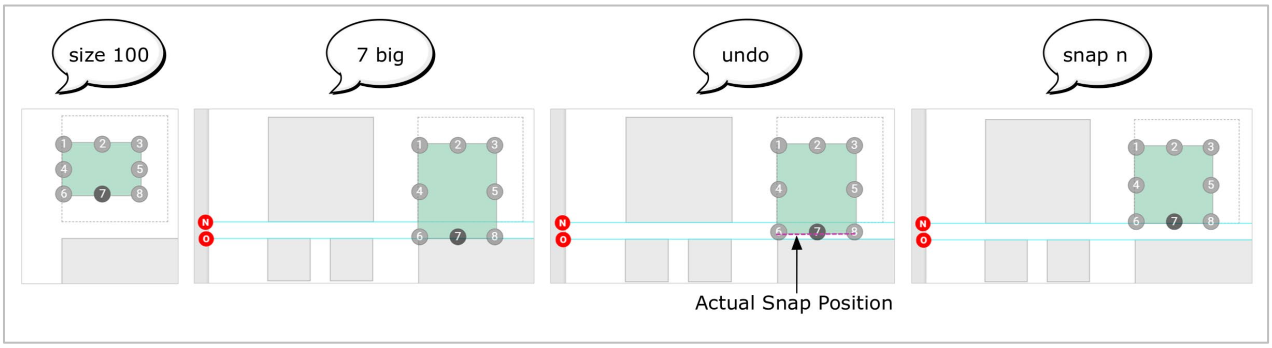


(a) **(b)** **(c)**

Figure 4.7. UserSnap: (a) transformation size is adjusted to 120 pixels using speech command "size 120", (b) object is resized from left-side by giving command "4 big", resulting in two guidelines "D" and "E" (within 100 pixels threshold from left side of the object) being displayed (c) "snap e" command is then issued to snap the object to desired guideline

4.3.1.3 AutoSnap

This approach combines UserSnap with an additional automatic snapping feature where objects automatically resize to the closest available snap location (within a 100 pixel threshold). An "undo" voice command is also available to address scenarios where users do not require an automated snap – this results in the object being returned to its original size prior to the automatic snapping action (Figure 4.8). Users can then still utilise the "snap x" command (similar to UserSnap) to adjust the object's size to any available snap points. A potential advantage of AutoSnap is that it can make snapping actions more efficient by reducing the need for users to always have to state a vocal command to perform a snap (which is required in UserSnap). However, there is also the potential within AutoSnap for undesired resizing actions which could be tedious and frustrating for users, whereas UserSnap provides full control over whether to perform a snapping action. It was therefore decided to investigate whether UserSnap and AutoSnap present any benefits when adjusting an object's dimensions and whether users have a preference for a particular technique.



(a) (b) (c) (d)

Figure 4.8. AutoSnap: (a) the transformation size value is set using command "size 100", (b) object is resized from bottom using "7 big" command, which results in two guidelines "N" and "O" becoming visible (within 100 pixels threshold) and the object is automatically snapped to closest available snap guideline "O", (c) an "undo" command is issued to return object to actual resize position, (d) finally "snap n" command is issued to snap the object at target snap guideline

4.3.2 Methodology (User Study 2)

User evaluations were conducted online with the involvement of people with physical impairments to explore the efficacy of the speech-controlled object resize approaches (i.e., NoSnap, UserSnap, and AutoSnap).

4.3.2.1 Participants

Twenty-five participants with physical impairments (17 male, 8 female) were recruited through online advertisements (via social media such as Facebook, Slack, and LinkedIn), as well as from existing network links. Participants were aged between 23-50 years (M=33.76, SD=6.94) and all were fluent English speakers. Participants were also requested to self-assess their experience with interface prototyping, graphical manipulation, and speech technology (these details are provided in Table 4.1).

Table 4.1 Participants information: Physical impairments and condition details; IP = Interface Prototyping (Software); GM = Graphical Manipulation (Software); ST = Speech Technology; AT = Assistive Tools

ID	Age/ Gender	Physical Impairments	Condition Details	Technical Experience
P1	42 (F)	Repetitive Strain Injury (RSI) Since (2016)	Difficulty in using fingers; Wrist Pain occasionally; Sometimes joint swelling and stiffness;	IP: Expert; GM: Expert; ST: Dragon software, Apple Siri; AT: Vertical mouse, Jellybean switch.
P2	35 (M)	RSI (Since 2010)	Hand tremors; Shooting pain in hands and arms; Pain in wrists; Tingling sensation in fingers.	IP: Average; GM: Average; ST: Dragon software, Google Assistant; AT: N/A.
P3	28 (M)	Tenosynovitis (Since 2020)	Wrist Pain; Joint swelling and stiffness; Difficulty in using fingers.	IP: Average; GM: Average; ST: Dragon; Apple Siri; AT: Head Tracker, Foot pedal.
P4	50 (F)	RSI (Since 2014)	Fatigue; Sore wrists occasionally; Shoulder pain; Pulsing pain in fingers.	GD: Average; IP: Average; ST: Google voice search services; AT: N/A.
P5	47 (M)	Motor Neuron Disease (Since 2016)	Muscle's weakness; Fatigue; Lack of balance; Unable to use hands	IP: Average; GM: Average; ST: Dragon software; Google Assistant AT: Tobii eye tracker.
P6	34 (M)	Muscular Myopathy (Since 2009)	Difficulty with walking without stick; Muscle's weakness; Fatigue; Lack of balance.	IP: Average; GM: Average; ST: Google speech services; AT: NA.
P7	26 (M)	Multiple Sclerosis (Since 2017)	Problem with balance; Tiredness; Numbness in fingers.	IP: Expert; GM: Expert; ST: Windows speech recognition; Google speech services, AT: NA.
P8	30 (M)	Tendinitis (Since 2015)	Fatigue; Pinched nerve; Muscle strains; Difficulty in holding stuff.	IP: Expert; GM: Expert; ST: Talon Voice, Google voice search; AT: Eye tracker.
P9	29 (F)	RSI (Since 2016)	Wrist pain, Pain in shoulders and upper arms; Tiredness; Stiffness in joints.	IP: Average; GM: Average; ST: Google Assistant, Samsung Bixby; AT: Head Tracker, USB Triple Foot Switch Pedal
P10	37 (M)	Lost Limb (Since 2018)	Amputated right arm	IP: Expert; GM: Expert; ST: Dragon software; AT: Foot pedal.
P11	36 (F)	RSI (Since 2012)	Weakness; Throbbing pain effect on hands occasionally; shoulders pain; Sometimes joint swelling at wrist.	IP: Average; GM: Average; ST: Apple Siri, Google voice search; AT: Jellybean Switch.
P12	29 (M)	RSI (Since 2015)	Discomfort in hands; Pain in fingers; Tiredness in arms;	IP: Average; GM: Expert; ST: Dragon software, Amazon Alexa; AT: NA.
P13	30 (M)	Spinal Muscular Atrophy (Type 2) (Since 1999)	Uses powered chair; Cannot walk since age 3; Unable to move hands and legs; Muscle's weakness; Lack of balance.	IP: Average; GM: Average; ST: Talon voice, Dragon software, Google Assistant; AT: Eye tracker, Head Pointer.
P14	23 (M)	RSI (Since 2017)	Shooting pain in hands and arms; Hand tremors occasionally; Tingling; Pain in wrists.	IP: Average; GM: Average; ST: Google speech services; AT: NA.
P15	26 (F)	RSI (Since 2018)	Aching fingers; weakness in hands and arms muscles; Numbness in fingers; painful wrists.	IP: Average; GM: Expert; ST: Mac voice control, Google Assistant; AT: NA.
P16	33 (M)	Multiple Sclerosis (Since 2012)	Fatigues; Numbness in arms and legs; Clumsiness; Lack of balance.	IP: Average; GM: Average; ST: Amazon Alexa, Apple Siri; AT: NA.
P17	42 (F)	Motor Neuron	Uses walking stick; Arms and	IP: Average; GM: Average;

		Disease (MND) (Since 2017)	shoulders pain; Fatigue.	ST: Windows speech recognition, Google speech services; AT: Head Tracker, Eye Tracker.
P18	29 (M)	RSI (Since 2017)	Hand tremors; Shooting pain in hands and arms; Pain in wrists; Muscle weakness.	IP: Average; GM: Expert; ST: Google Home, Dragon software; AT: NA.
P19	35 (M)	RSI (Since 2006)	Shoulder pain; Tiredness in forearms; Sore wrists occasionally; Pulsing pain in fingers.	IP: Expert; GM: Expert; ST: Google speech services, Mac voice control; AT: Foot pedal.
P20	40 (F)	Motor Neuron Disease MND (Since 2018)	Weak grip, Hard to climb stairs, Weak muscles	IP: Average; GM: Expert; ST: Google speech services; AT: NA
P21	28 (M)	Shoulder Impingement Syndrome (2020)	Weakness in arms, Pain in shoulders, Severe pain when lift arms above head	IP: Average; GM: Average; ST: Dragon software, Apple Siri; AT: NA.
P22	38 (M)	RSI (Since 2018)	Pain in forearms and elbows; Throbbing sensation in fingers; joint swelling sometimes.	IP: Average; GM: Expert; ST: Windows speech recognition; AT: NA.
P23	34 (F)	RSI (Since 2011)	Occasionally severe pain in hands; Tiredness in shoulders and upper arms;	IP: Average; GM: Average; ST: Samsung Bixby, Google Assistant; AT: NA.
P24	23 (M)	RSI (Since 2017)	Stiffness of joints; feeling of numbness in fingers; muscles weakness;	IP: Expert; GM: Expert; ST: Dragon, Google Assistant; AT: Trackball mouse.
P25	40 (M)	RSI (Since 2005)	Fatigue; Shoulder pain; sore wrist; throbbing pain in hands and fingers	IP: Average; GM: Expert; ST: Dragon, Talon voice; Apple Siri; AT: NA.

4.3.2.2 Apparatus

All participants used their own computer and microphone for voice input, as well as their own chosen form of switch for enabling the speech recogniser. Fifteen participants utilised a keyboard (i.e., the spacebar key), seven used dragon software Dragon (2023) (e.g., via a vocal command such as ‘press spacebar’), two utilised a foot pedal, and one used a Jellybean switch. The Google Chrome browser (version 80 or above) was required for experimental tasks to ensure browser compatibility with the Web Speech API. Testing sessions were conducted online via Zoom and Microsoft Teams video conferencing platforms.

4.3.2.3 Procedure

Ethical approval was obtained from Birmingham City University’s ethical committee to conduct this study (Appendix 8.2: Object Resize [page 167]). A recruitment advertisement was posted through social media networks (via Facebook, Slack community, Twitter, and LinkedIn) requesting those interested to contact the named researcher. A testing session was then scheduled for participants using their preferred choice of online platform (e.g., Teams or Zoom). During the testing session, the researcher initially provided participants with a link to the research prototype which they were asked to access and then share their screen content. After an overview of the study was then given by the researcher, participants then went through the PIS (participant information sheet) and were required to read and complete a consent form (via selecting a confirmation button within the prototype). Participants then completed the pre-test questions requesting details around demographic information and technical experience in relation to interface prototyping applications and speech technology (Appendix 8.2: Object Resize [page 170]). Participants were also asked about the nature of their impairments and any assistive tools they utilise. Once the survey had been completed, participants were asked to complete a training task which involved resizing a small interaction object to a larger size (highlighted through a target size placeholder) using relevant voice commands. An additional reference object was also provided for the UserSnap and AutoSnap practice tasks to ensure participants were able to familiarise themselves with the object snapping features.

After completion of the training task participants moved onto the main tasks for the first interaction approach they had been assigned to use (conditions and task order were counterbalanced to minimise order bias). There were ten object resize tasks for each interaction approach (i.e., 30 tasks in total) that involved adjusting the size of a green colored interaction object to the dimensions of a target placeholder (displayed as blank dotted box). At the start of a task, the interaction object was placed at the center of its corresponding target size placeholder to ensure that all sides of the object had to be manipulated in size (the position of the objects on the canvas could not be altered). A variety of interaction object and placeholder sizes were selected to ensure participants perform a range of different resizing tasks (Figure 4.9). Similar to the exploratory study, these were informed though an analysis of standard interface elements within mainstream social media applications (i.e., in terms of profile covers, thumbnails, icons, logos, etc.).

To initiate a task, participants activated the speech recogniser using their chosen form of switch input and then started to resize interaction objects as accurately as possible via the available speech commands and features. Participants continued with a task until they felt the interaction object's dimensions had been accurately adjusted to the corresponding target placeholder's size. The same process was repeated until all ten tasks for the condition had been attempted – a SUS form was then administered for participants to complete. After the same process had been completed for all three conditions, a semi-structured interview was conducted where participants were asked about what they liked and disliked about each interaction approach, their preferred method, overall impressions, and any suggestions for improvement. The testing session with each participant was video-recorded for later analysis. All testing sessions lasted between 50 minutes to 1 hour.

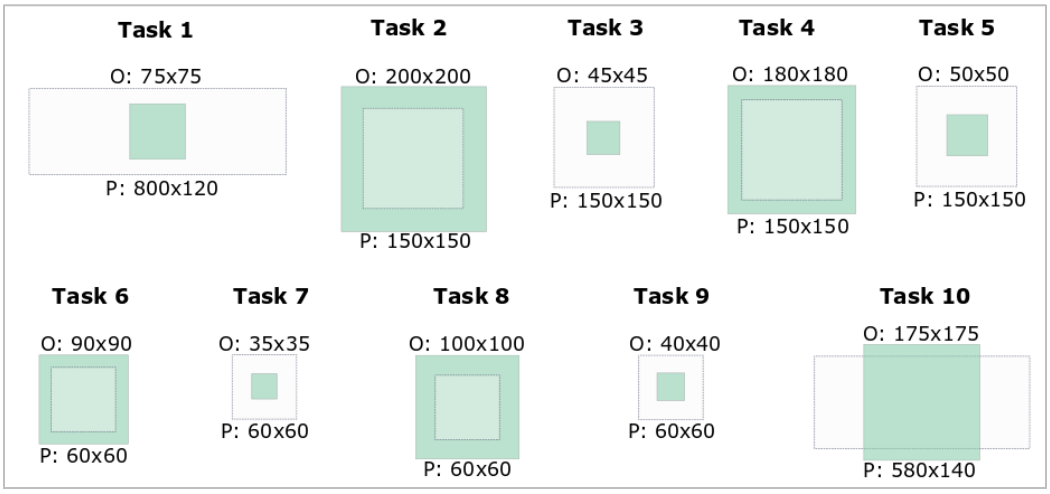


Figure 4.9. Interaction Object (O) and Placeholder sizes (P) for experimental tasks

4.3.2.4 Measures

The measurements included task completion time, resize accuracy, speech recognition performance, and SUS scores to evaluate the three interaction approaches. Task completion time was measured in milliseconds across each task screen which was captured from when a participant clicked the 'start task' button before starting a task until they then selected the 'next task' button present in the sidebar upon completion of a task. Resize accuracy was again measured through calculating the Euclidean distance between the final size of interaction objects and target placeholder dimensions (using the same approach outlined in Section 4.2.3.4). Speech recognition performance was measured via the total number of speech commands issued, as well as speech recognition errors which were again classified into same three different categories highlighted in Section 4.2.3.4: 'Speech Misrecognition', 'System Error', and 'Unsupported Commands'. SUS was used to evaluate perceptions of usability for each interaction approach.

4.3.3 Results

The Shapiro-Wilk's normality test (Shapiro and Wilk, 1965) found task completion, resize accuracy, speech recognition performance, and SUS scores were not normally distributed, whilst data related to the total number of speech commands was found to be normally distributed. A non-parametric Friedman test of differences for repeated measures (Zimmerman and Zumbo, 1993) was used with Bonferroni correction to perform the analysis. Wilcoxon signed rank (Woolson, 2008) was used as a post-hoc test to further analyse the differences in task completion time, resize accuracy, speech performance, and SUS scores. A one-way repeated measure ANOVA was utilised to analyse the total number of speech commands for each interaction approach.

4.3.3.1 Task Completion Time

The mean task completion time for NoSnap was 12.11 minutes (SD=1.71), UserSnap 12.64 minutes (SD=1.69), and AutoSnap 9.09 minutes (SD=1.03). Friedman test results found significant differences in task completion time (X2=0.001, df=2, p<0.05). Post-hoc Wilcoxon signed rank also showed a significant difference in task completion time between NoSnap and AutoSnap (Z=-4.37, p<0.001) and between UserSnap and AutoSnap (Z=-4.37, p<0.001). No significant differences were found between NoSnap and UserSnap (Z=-1.87, p=0.061) (Figure 4.10).

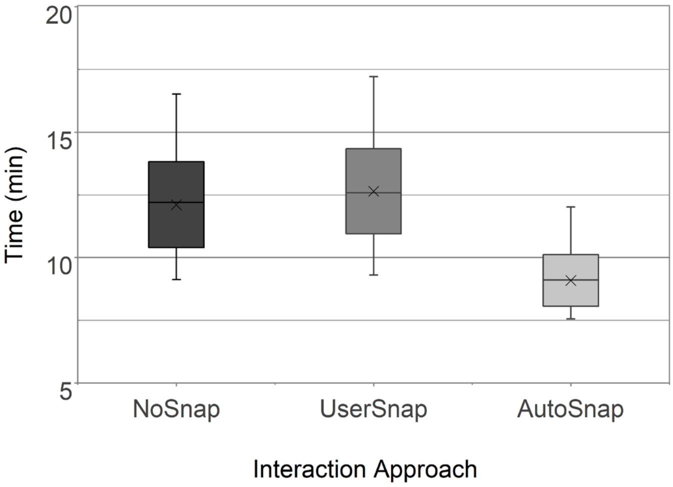


Figure 4.10 Average Task Completion Time

4.3.3.2 Resize Accuracy

The mean resize accuracy calculated from Euclidean distance values for NoSnap was 0.84 pixels (SD=0.81), 0.32 pixels (SD=0.47) for UserSnap, and 0.24 pixels (SD=0.46) for AutoSnap. Friedman test results showed significant differences in resize accuracies between all interaction approaches (X^2=0.001, *df*=2, *p*<0.05). The post-hoc Wilcoxon signed rank found a significant difference in resize accuracy between AutoSnap and NoSnap (*Z*=-8.39, *p*<0.001), and also UserSnap and NoSnap (*Z*=-7.45, *p*<0.001). However, no significant differences were found between AutoSnap and UserSnap (*Z*= -1.81, *p*=0.70) (Figure 4.11).

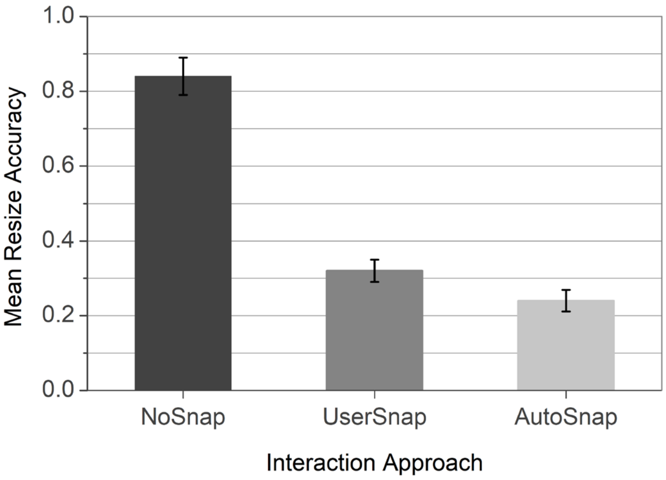


Figure 4.11 Mean resize accuracies (via average Euclidean distances)

4.3.3.3 Speech Performance

The total number of vocal commands issued across all 25 participants for NoSnap was 3591 (SD=11.31), 4072 (SD=10.06) for UserSnap, and 2986 (SD=10.29) for AutoSnap (Figure 4.12). A statistically significant difference was observed between NoSnap, UserSnap, and AutoSnap in relation to total number of speech commands (*F* (2, 48) = 107.40, *p*< 0.001, partial *η2* = 0.817). The post hoc Least Significant Difference (LSD) test also highlighted a significant difference between NoSnap and UserSnap (sig = 0.001, *p*< 0.05), UserSnap and AutoSnap (sig = 0.001, p<0.05), and also NoSnap and AutoSnap (sig = 0.012, p< 0.05).

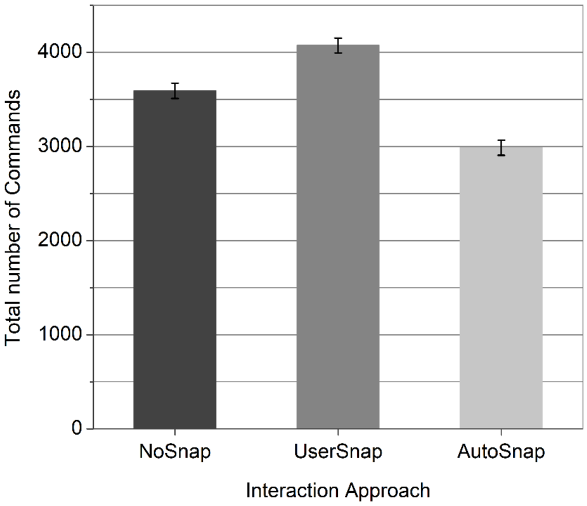


Figure 4.12 Total number of speech commands

There were 182 (5.07%) ‘Speech Misrecognition’ errors in NoSnap, 219 (5.38%) for UserSnap, and 150 (5.02%) for AutoSnap. Friedman test results found no statistically significant differences for ‘Speech Misrecognition’ across all three approaches (X^2=0.31, *df*=2, and *p*>0.05). For ‘System Errors’, 73 (2.03%) speech commands were related to NoSnap, 87 (2.14%) in UserSnap, and 56 (1.87%) in AutoSnap. Friedman test results found no statistically significant differences for ‘System Errors’ across all three approaches (X^2=0.20, *df*=2, and *p*>0.05). There were also 6 (0.17%) ‘Unsupported Commands’ issued across NoSnap, 11 (0.27%) for UserSnap, and 9 (0.30%) for AutoSnap. Friedman test results did not find statistically significant differences related to ‘Unsupported Commands’ across all three approaches (X^2=0.30, *df*=2, and *p*>0.05).

4.3.3.4 Usability Scores

The mean SUS score for NoSnap was 70.40 (SD=3.72), 73.20 (SD=11.42) for UserSnap, and 82.00 (SD=3.75) for AutoSnap. NoSnap and UserSnap scores can therefore be labelled as exhibiting a “Good” level of usability, while AutoSnap can be labelled as “Excellent” (Bangor et al., 2009). Significant differences were found across the three interaction approaches using Friedman test (X2=0.001, *df*=2, *p*<0.001). The post-hoc Wilcoxon signed rank also highlighted a significant difference between NoSnap and AutoSnap (*Z*=-4.30, *p*<0.001), NoSnap and UserSnap (*Z*=-2.10, *p*<0.001), as well as AutoSnap and UserSnap (*Z*=-3.43, *p*<0.001) (Figure 4.13).

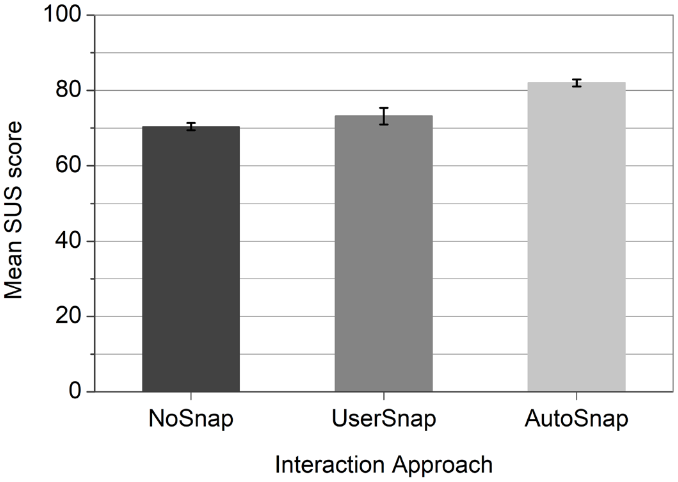


Figure 4.13 SUS score across three speech-controlled object resize approaches

4.3.3.5 Qualitative Feedback

User Perceptions of Voice Snapping Techniques:

All participants provided positive feedback in terms of using the three different object resizing techniques. Sixteen participants preferred AutoSnap, seven preferred UserSnap, and two preferred NoSnap. Positive comments in relation to AutoSnap highlighted that it was ‘intuitive’, ‘reliable’, ‘saves time’, and ‘reduces the use of voice commands’ over the other methods. In particular, thirteen participants stated that the reduction of commands in AutoSnap could help to avoid vocal strain in comparison to UserSnap (where snapping commands always have to be issued).

“… if I am working on a real project with complex design, then easier activities like resize can really be facilitated using Auto Snap and it will save a lot of effort and time”. (Participant 22)

P16 stated that AutoSnap is more helpful for resizing square shaped objects with matching dimensions (e.g., 50x50 or 100x100), although it is not as efficient as UserSnap for objects where width and height may differ (e.g., 800x120, 400x200, etc.). For instance, when using corner transformation handles (i.e., 1, 3, 6, 8) in AutoSnap for objects with different dimensions (i.e., 800x120), two snapping actions could occur to both horizontal and vertical snap points simultaneously (which may be undesired). Seven participants therefore directly stated that UserSnap was more effective in this scenario as it provided more control over resizing actions (“*I like UserSnap as it works consistently for resizing object from all sides so I felt more control over transformation when using this method*”) (Participant 12). Two participants who preferred NoSnap highlighted that they liked the simplicity of the approach, and that it was straightforward and easy to understand.

Speech Commands:

There were no major speech recognition issues identified and all participants were able to successfully complete all object resize tasks. It was observed that there was a range of misrecognised voice commands such as the “small” command being identified as different homophonic words including “mole” or “mall”. Furthermore, it was found that three participants used directional commands on occasions to resize the dimensions of objects such as “left big” instead of “4 big” (to make an object bigger from the left boundary of the object). Similarly, “go down” instead of “2 small” was used to make the object smaller from the top boundary and “right right” instead of “5 big” to make object bigger from right object boundary. One participant also used a “5 large” command as opposed to using “5 big” and “left shorter” for “4 small”. Participants were asked about the motivation behind using directional commands in the semi-structured interview where they highlighted that this was an unintentional mistake (as opposed to not understanding the task requirements). Moreover, two participants attempted to chain commands together using a combination of the size command coupled with transformation size (e.g., “4 big 10”) with an expectation that it would complete two actions (“4 big” and “size 10”). These participants highlighted that they wanted to avoid issuing a separate command for changing the size value when resizing the object and thought combining size value with resize command “big” would save effort and complete both actions using single command. Since this study was primarily focused around the use of short vocal commands, further research is required to explore whether the use of chained commands can help users in efficiently completing a task or whether this presents any usability challenges. Although participants had a range of technical and design experiences however, no participants highlighted significant issues related to the use of presented speech commands and in completing evaluation tasks.

Transformation Size Estimation and use of the Ruler:

In relation to NoSnap, eight participants emphasised an issue around estimating the size of transformations to efficiently complete a resize task – e.g., “... *It is hard to estimate correct transformation size value in first attempt so I tried to use ruler but then I had to calculate distance between ruler values to get correct transformation size value which required effort*”. (Participant 15)

“… ruler helped to get bigger transformation size value … but it is hard to estimate correct value when you need to make small adjustments” (Participant 11).

This was a common theme with NoSnap, although it was not highlighted in relation to UserSnap and AutoSnap as a significant challenge. It was observed that twelve participants attempted to use the ruler in the NoSnap approach and also highlighted that they think it might be useful for this approach:

“I used ruler a little bit in the beginning in NoSnap and it was good exercise to understand transformation size values but later I was able to figure out myself what 20 pixels or 50 pixels values meant without relying on it”. (Participant 13)

Participant 12 suggested that they would like to use the ruler if they were able to see the ruler values closer to the interaction object, so that they do not have to regularly adjust their gaze to the top and left sides of the interface (i.e., where the ruler is located). A similar point was raised by another participant: “... *ruler should only be displayed when user needs to estimate the size value. Currently ruler and snap guidelines are displayed at a time on the* *canvas, it should be optional to use ruler and when it is not required there should be 'hide' command option*" (Participant 8).

4.4 Conclusions

This chapter has addressed the research question highlighted in Section 4.2.1 by developing new voice-controlled resizing techniques for manipulating the size of graphical objects. An initial exploratory study demonstrated that the use of object handles (typically used in mainstream applications) can be tailored for voice control and presents a viable interaction approach. Participants with physical impairments also found all three resizing approaches (utilising a mixture of handles and snapping tools) presented in the second study to be viable and usable, although AutoSnap was perceived to be more efficient, accurate, and usable than both NoSnap and UserSnap. Subjective feedback from participants also correlated with quantitative findings with participants providing positive comments around the efficiency and intuitive nature of the snapping approaches (over NoSnap). The study considered users with a diverse range of technical and design experiences while no significant challenges were faced across all participants therefore presented speech-controlled approaches can support users with a diverse range of technical and graphic design experiences.

Moreover, the clear preference for the UserSnap and AutoSnap techniques highlights that the snapping features developed were beneficial in the context of the object manipulation tasks that participants completed. This work therefore contributes a deeper understanding around the feasibility of voice-controlled snapping approaches to support people with physical impairments when completing digital creative tasks. In particular, the results indicate that common snapping techniques operated via traditional input devices (Baudisch et al., 2005; Bier, 1986; Masui, 2001) can be effectively tailored for speech interaction (within a creative context), thus building on initial work in this area (Van der Kamp and Sundstedt, 2011).

Whilst the results highlighted an overall benefit for AutoSnap, there is still the possibility that UserSnap can be a more efficient and effective approach in some scenarios. This was highlighted through feedback from participants who felt that UserSnap was more appropriate when looking to adjust the corner transformation handles on objects (to avoid potential undesired automatic snapping to both vertical and horizontal snap points). This will also likely be the case in scenarios where multiple snap points are located in close proximity to each other – automatic snapping here may well lead to user frustration as it increases the likelihood that objects will snap to the incorrect location. Conversely, UserSnap may present benefits here as it would provide users with full control over which snap point they wish to target. A hybrid approach utilising some degree of user control and automation may likely be optimal in certain scenarios, although additional research is required to empirically investigate this further and understand the nuances around object snapping via speech control. Future work also needs to explore potential adaptations to the voice commands used – for instance, participants used directional commands on occasions (e.g., “left big”), as well as chaining different commands together (e.g., “4 big 10”). The current system did not support these types of commands, hence further research around these areas could inform and enhance the usability of the existing approaches developed. Moreover, related research has previously explored the potential of vocal commands to augment the workflow of non-disabled professional designers alongside traditional input devices (i.e., a mouse, keyboard, stylus) (Kim et al., 2019; Laput et al., 2013; Sedivy and Johnson, 1999; Srinivasan et al., 2019). AutoSnap and UserSnap may therefore also have wider potential to enhance the creative flow of non-disabled designers, although further work is required to confirm whether this may present interaction benefits.

5 STUDY 3: OBJECT ROTATION USING SPEECH INTERACTION

5.1 Introduction

This chapter presents research investigating object rotation via speech interaction to support design workflows for people with physical impairments. Similar to object positioning and resizing, rotation is essential in object manipulation activities associated with digital design work. In mainstream creative design applications (such as Photoshop and Illustrator) users can rotate the objects through dragging transformation handles via a mouse or by inserting a numerical value of required transformation angle using the properties panel. While this is a common approach that is widely used across different design software, it requires the use of dragging movements via mouse or using keyboard hence presenting significant challenges for people with physical impairments in rotating the objects on a design canvas. Previous literature has investigated object rotation using multimodal input approaches with speech interaction – for example, Williams and Ortega (2020) explored the use of voice interaction (e.g., via commands such as “spin”, “roll”, “yaw”, etc.) in combination with hand gestures for manipulating 3D objects. Similarly, Alibay et al. (2017) utilised the combination of speech and gestures for the selection and rotation of objects within a 3D digital design context. Furthermore, House et al. (2009) explored the use of speech for rotating objects with the help of a robotic arm using non-verbal vowel sounds (e.g., “iy” and ”ae” for left and right movements). However, whilst these studies demonstrated the potential of voice control to support object rotation, there has been a lack of empirical work investigating the rotation of objects using speech commands as the primary input modality within a creative 2D visual design context. The research presented in this chapter addresses the limited work in this area by three key research studies investigating how object rotation can be facilitated through speech interaction.

Whilst previous literature has investigated speech-controlled object positioning and resizing approaches (which helped to inform the commands used in the studies conducted in Chapters 4 and 5), no previous studies have explicitly explored object rotation approaches using voice interaction. This elicitation study was therefore an important initial step to better understand which types of vocal commands would be most approach to support people with physical impairments in rotating objects on a design canvas. The findings from the elicitation study informed a further interactive exploratory investigation with users who have physical impairments to explore their perceptions of rotating different design assets using the voice commands highlighted by participants. The results from this study identified some interaction challenges (e.g., issues in estimating the size of rotation transformations) which motivated the development and evaluation of a final research prototype comprised of three speech-controlled object rotation techniques. Finally, a user evaluation with participants who have physical impairments is then presented to evaluate the three object rotation approaches (Baseline, Fixed-Jumps, and Animation) with results demonstrating a clear preference for the Animation technique. The findings from the studies conducted in Chapters 3 and 4 influenced the design of these approaches through common design points such as the selection of short vocal commands and the ability to manipulate objects via specific transformation sizes. However, given that rotation is a fundamentally different form of object manipulation (compared to positioning and resizing), it was felt that the interaction techniques developed (in Chapters 3 and 4) would not directly align for rotation transformations (thus leading to the development of new tailored voice-controlled rotation approaches). This chapter therefore addresses the limited work in this area through presenting four primary contributions: (1) results from an elicitation study highlighting the vocal rotation commands people with physical impairments would prefer to use; (2) an exploratory study highlighting the viability of rotating objects via voice control and the unique challenges this can present; (3) the development of three new speech interaction approaches for rotating digital assets; (4) a user evaluation with people who have physical impairments presenting new insights around the use of speech interaction for object rotation.

5.2 Research Question

The following overarching research question was developed to address the lack of work in this area:

RQ: How can voice interaction support rotation transformations of digital objects?

In alignment with the other studies conducted in Chapters 3 and 4, this question again encouraged an exploratory approach to support better understanding around the feasibility of object rotation via voice control and the potential of new approaches to support effective transformation of graphical objects.

5.3 Elicitation Study – Voice Commands for Object Rotation

Given the limited work around rotating objects via speech interaction within a 2D design space, it was unclear which types of commands would be most suitable to support this form of object transformation. Understanding users' thinking and observing their actions is essential to identify intuitive and appropriate methods prior to creating interactive systems (Wobbrock et al., 2009). It was therefore important to conduct an initial elicitation study with users who have physical impairments to determine the vocal commands that could be associated with object rotation (in relation to creative design work). To facilitate this work, a research prototype was designed consisting of four different types of design assets (e.g., images, text, lines, and icons) where users had to verbalise the commands they would use to rotate the objects to relevant target placeholders. Each type of object involved four tasks where two objects were rotated in a clockwise direction and another two were rotated in anti-clockwise directions. Furthermore, two tasks for each type of object involved rotation with "smaller" transformations and two with "larger" transformations (based on the angle of target placeholders). Large transformations were considered as having a rotation angle of 180 degrees or above, while smaller transformations included rotation manipulations of less than or equal to 50 degrees. These larger and smaller threshold transformations were informed by earlier research investigating rotation techniques where they considered angles up to 180 as large transformations and angles between 20 to 60 degrees as small transformations (Laviola and Katzourin, 2007; Poupyrev et al., 2000). The variety of objects and range of transformations were presented as referents to ensure participants performed a variety of activities whilst verbalising vocal commands.

5.3.1 Methodology

5.3.1.1 Participants

Twelve participants with physical impairments (8 male and 4 female) were recruited through online advertisements via social media (Facebook, Slack, LinkedIn, and Twitter). Participants were aged between 21 to 52 years (M=35.25, SD=7.82) and all were native English speakers. Participants provided demographic information and details around the nature of their physical impairments, as well as details around their experience with interface prototyping, graphical manipulation applications, speech technology, and assistive tools. Eight participants identified as experiencing RSI, two with motor neurone disease (MND), one with multiple sclerosis (MS), and one with tenosynovitis (Table 5.1). Six participants had average experience with graphical manipulation software, whilst six identified has having expert level experience. In terms of interface prototyping applications, seven participants were identified as having average experience while five had expert level experience. Further details about participants' experience with speech technologies and use of other assistive technology tools is provided in Table 5.1.

Table 5.1 Participants Details: Age, Gender, Physical Impairment and Condition Details, GM = Graphical Manipulation (Software), IP = Interface Prototyping (Software), ST = Speech Technology, and AT = Assistive Technology

ID	Age/ Gender	Physical Impairments	Condition Details	Technical Experience
1	35 (M)	Repetitive Strain Injury (RSI) (Since 2020)	Fatigue; Shooting pain in hands and arms; Aching fingers.	IP: Average; GM: Average; ST: Dragon software, Google Assistant; AT: N/A.
2	42 (F)	RSI (Since 2017)	Clumsiness; Forearms pain; Numbness in hands and fingers.	IP: Expert; GM: Expert; ST: Dragon software, Apple Siri; Google Home AT: N/A.
3	26 (M)	Multiple Sclerosis (Since 2019)	Lack of balance; Difficulty with walking; Fatigue; Numbness and tingling sensation in hands.	IP: Expert; GM: Average; ST: Amazon Echo; AT: Eye Tracker.
4	35 (M)	RSI (Since 2018)	Tiredness; Muscle cramps in forearms; Pins and needles; Throbbing pain in hands;	GD: Average; IP: Average; ST: Google Assistant; AT: Mechanical Switch.
5	37 (M)	Tendinitis (Since 2021)	Stiffness; Joints swelling; Hands and wrist pain; Difficulty with holding objects.	IP: Average; GM: Average; ST: Talon Voice; AT: N/A.
6	47 (M)	Motor Neurone Disease (MND) (Since 2012)	Weak grip; Weakness in muscles; Harder to climb stairs; Muscle twitching.	IP: Average; GM: Expert; ST: Dragon software; AT: Eye tracker; Head pointer.
7	29 (F)	RSI (Since 2019)	Wrist pain; Pain in shoulders and neck; Fatigue; Tiredness.	IP: Expert; GM: Average; ST: Talon Voice, Google voice search; AT: Foot pedal.
8	22 (M)	RSI (Since 2021)	Difficulty in holding stuff; Painful wrists; Numbness in fingers; Pain in forearms.	IP: Expert; GM: Average; ST: Windows speech recognition; AT: N/A.

9	30 (M)	RSI (Since 2019)	Frequently feeling tired; Fatigue; Pain in shoulders and forearms; Tingling sensation in hands.	IP: Expert; GM: Expert; ST: Google Assistant, Apple Siri; AT: N/A
10	35 (M)	RSI (Since 2017)	Severe pain in hands occassionlly; Sore wrists; Pain in forearms and elbows.	IP: Average; GM: Expert; ST: Dragon software; Google Assistant; AT: Foot pedal.
11	36 (F)	RSI (Since 2014)	Pulsing pain in fingers; Difficult to move fingers; Stiffness; Frequent pain in shoulders and arms.	IP: Average; GM: Average; ST: Apple HomePod; AT: Wireless pen mouse.
12	50 (F)	MND (Since 2010)	Pain in arms and shoulders; Fatigue; Difficult to lift hands; Diffiulty in walking.	IP: Average; GM: Average; ST: Dragon software; AT: Eye tracker, Optikey.

5.3.1.2 Apparatus

The testing sessions were conducted remotely using Zoom or Microsoft Teams depending on each participant's preference. All participants used their own computer or laptop during the study. Adobe XD was used to create the prototype for the elicitation study (Figure 5.2).

5.3.1.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.3: Object Rotation [page 173]). The recruitment advertisement was posted on social media networks (via Facebook, Slack community, Twitter, and LinkedIn) where those interested were requested to contact the named researcher. A testing session was then scheduled at a suitable time on the participant's platform of choice (e.g., Teams or Zoom). During the evaluation session, the researcher shared their screen content where they displayed the study information sheet to participants, followed by the consent form where participants were requested to provide formal consent. Participants were then asked pre-test questions in relation to demographic information, technical experience of using graphical manipulation and interface prototyping applications, and speech technology. Participants were also asked about the nature of their physical impairment and any assistive tools they use to work with creative applications.

The Adobe XD prototype was then used to display the training task screen to participants, followed by the referents of the study presented as a set of tasks for which they had to verbalise the commands they considered most suitable. There was a total of sixteen actions which consisted of four different types of objects (i.e., images, text, lines, and icons – Figure 5.1). Each type of object involved four referents where objects were rotated in clockwise and anticlockwise directions, as well as requiring larger and smaller rotation transformations. The range of transformation angles and variety of objects were chosen to ensure participants could perform different rotation activities and then provide their vocal command suggestions. The four different object types were evaluated to observe whether these common design assets influenced the types of commands participants issued.

Participants were provided with a single training task before being presented with the referents for each object type. The training task consisted of a single object displayed on the blank canvas where participants were asked to verbalise the speech commands they would use to rotate the object to the target position (depicted as a semi-transparent copy of the object rotated to the target position). Once participants had completed the training activity, they progressed onto the main study tasks where a single object was presented on a canvas while a corresponding target placeholder was placed behind the object with a different rotation angle to the interaction object (Figure 5.2). The static rotation angle and *xy* positions of the interaction object were also displayed in the right sidebar of the interface.

Figure 5.1 Four different types of objects (image, text, line, icon) used for object rotation during Elicitation and Exploratory studies

Referents were then displayed in a counterbalanced order to minimise the order bias. Participants started by verbalising their initial observations and were encouraged to highlight the types of commands they would use to rotate the object to the required location. It is important to note that static images were presented to participants and that issuing commands did not alter the rotation angle of the object (the key motivation of the study was to extract potential vocal commands for rotation). After completing four tasks for an object type (e.g., images), they were encouraged to suggest any alternative commands other than the ones they had highlighted. They then started on the next set of four tasks for a different object type. At the end of testing session, participants were encouraged to provide general suggestions and feedback on the use of speech for rotating objects on a digital canvas. Sessions lasted between 20 to 25 minutes and were video recorded for later analysis.

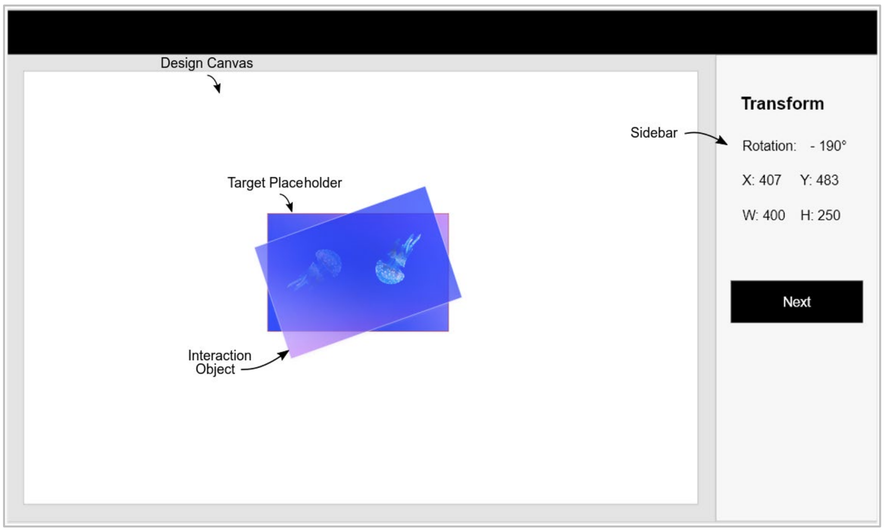


Figure 5.2 Elicitation study prototype interface showing creative assets (a design canvas, an interaction object, a target placeholder, and a sidebar consisting of attributes of the interaction object)

5.3.2 Results

Participants issued 225 voice command suggestions in total. The two most popular types of commands included “clockwise/anticlockwise” which was issued on 91 occasions (40.44%) and “left/right” commands which were stated 94 times (41.77%). Seven participants highlighted a preference for using “left/right” commands (across all object types and transformation sizes) while five participants had a preference for “clockwise/anticlockwise” – for example:

“… I would prefer to use ‘right’ and ‘left’ commands because these are short and easy to use words and for me this is regardless of any sort of shapes and objects being displayed”. (Participant 12)

It was also observed that object type (images, lines, text, icons) did not affect any users’ decision on which commands to use – for instance, if a user utilised clockwise/anticlockwise commands during the first set of tasks (e.g., images), they then utilised the same commands with the other types of objects (e.g., lines, text, or icons – unless asked to provide alternative suggestions at the end of the tasks). On occasions, participants also attempted to combine other words with the “left/right” and “clockwise/anticlockwise” commands. For instance, “turn clockwise/anticlockwise”, “rotate clockwise/anticlockwise”, “half clockwise/anticlockwise” “move clockwise/anticlockwise”, “spin right/left”, “rotate right/left”, “right up/down”, “left up/down”, “clockwise small/big”, and “rotate right twice”. A total of thirty-four alternative voice commands were also suggested (across all participants) when they were requested to provide these after completion of a set of tasks for an object type (Table 5.2).

In relation to objects with large transformations, a majority of participants suggested that they would repeat the same voice commands to get the object to the correct rotation position. For instance, if an object did not reach a target position after "left 180", then they would use another command (e.g., "left 20") to refine the final rotation position of the object. Three participants also suggested that they would use "flip" as a command for transforming objects that required larger rotation angles for example:

"I would use flip vertical and flip horizontal commands to rotate objects to reach at target place quickly and then I will adjust the target position using commands "set angle 10 and rotate left" (Participant 6).

Table 5.2 Elicitation study – Voice commands highlighted by each participant

ID	Frequently Used Commands	Other Suggested Commands
P1	Clockwise [xDegree], Anticlockwise [xDegree]	Torque right, Torque left, Forward, inverse.
P2	Clockwise, Anticlockwise	Counter-clockwise, Opposite x Deg,
P3	Right, left	Spin around x deg, spin over.
P4	Clockwise, anticlockwise	Left in, Left out, right down, left up, flip.
P5	Clockwise, Anticlockwise	Rotate towards down, rotate opposite up, mirror opposite.
P6	Right, Left	Flip vertical, flip horizontal, set angle x and rotate.
P7	Right, Left	Turn around, turn away, circulate towards, circulate, opposite.
P8	Right, Left	Revolve x deg, revolve around, revolve opposite, escalate x deg.
P9	Right, Left	Go straight, go x deg, go round, go around. Go opposite x deg.
P10	Clockwise, Anticlockwise	Rotate counter-clockwise, inwards x deg, outwards x deg.
P11	Right, Left	Flip over, flip, start rotate, stop rotate, rotate up, rotate down.
P12	Right, Left	Spin x deg, wheel x deg, rotate large, rotate small.

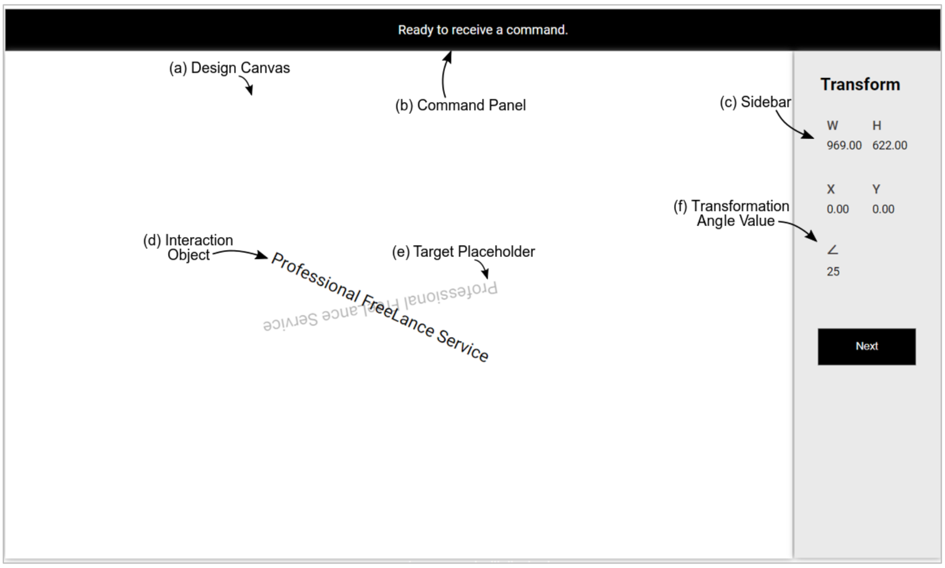
Participants did not highlight any issues related to objects with smaller transformation as they felt it would be easy to rotate these objects with a single or fewer attempts. A key theme emphasised by all participants is that they would prefer voice commands that are short and require less effort in pronouncing (e.g., to avoid vocal discomfort). Participant 8 suggested they would also consider a custom set of commands (as supported by Dragon software) comprised of "random" commands that are short and easy to pronounce (e.g., "alpha" or "bravo" instead of "right").

5.3.3 Elicitation Study Discussion

Based on the findings and suggestions from the elicitation study, it was identified that most participants preferred to use “left/right” commands for object rotation tasks (followed by “clockwise/anticlockwise). Moreover, object type (i.e., images, text, lines, and icons) and transformation size did not affect decisions around choice of commands. It was therefore decided that since there was a preference for “left/right” and that participants found these short and easy to pronounce, these would be taken forward for an exploratory study investigating users’ perceptions of the commands to interactively manipulate the rotation angle of digital assets.

5.4 Exploratory Study

An exploratory study was conducted with 12 participants who have physical impairments to evaluate perceptions around rotating graphical assets using the commands identified through the elicitation study. A web-based research prototype was developed using HTML, CSS, and JavaScript, along with the Web Speech API for speech recognition (Figure 5.3). The prototype included a design canvas containing an interaction object along with a semi-transparent target placeholder highlighting the target rotation orientation. The prototype also contained a black header where voice commands issued by users were displayed, as well as a sidebar (on the right side) where standard properties of the interaction object were presented (i.e., width, height, *xy* positions, and the current rotation angle).



Ready to receive a command.

Transform

W 969.00 H 622.00

X 0.00 Y 0.00

∠ 25

Figure 5.3 Exploratory study research prototype consisting of (a) design canvas, (b) command panel, (c) sidebar, (d) interaction object, (e) target placeholder, and (f) transformation angle value

A single interaction object (image, text, line or an icon) was displayed on the design canvas and could be rotated using either “left” or “right” speech commands combined with a transformation value (e.g., “left 10” or “right 20”). Switch input (e.g., a keyboard, mechanical switch, head tracker, or a foot pedal, etc.) could also be utilised to initiate the speech recogniser before starting a rotation task. Audio feedback via a popping sound effect was played after a voice command had been issued to make users aware that their input had been recognised.

5.4.1 Methodology

5.4.1.1 Participants

Twelve participants with physical impairments (6 female and 6 male) were recruited through online advertisements using social media networks such as Facebook, Twitter, and LinkedIn. Participants were aged between 22 to 50 years (M=33.75, SD=7.47) and all were native English speakers. Table 5.3 details participants’ demographics, nature of physical impairments, and experience with interface prototyping, creative applications, speech technology, and assistive tools. Participants in this exploratory study were different from those who took part in the first study to eliminate any potential bias being introduced.

Table 5.3 Participants Details: Age, Physical impairment, Condition Details, and Technical Experience Details.

ID	Age/ Gender	Physical Impairments	Condition Details	Technical Experience
1	42 (M)	Repetitive Strain Injury (RSI) Since (2013)	Pain in hands and arms; Pain in wrists; Pins and Needles feel in fingers.	IP: Expert; GM: Average; ST: Dragon software, Apple Siri; AT: Eye tracker.
2	35 (F)	RSI (Since 2011)	Fatigue; Pain and Numbness in fingers; Wrist Pain;	IP: Average; GM: Expert; ST: Dragon software, Google Assistant; AT: Foot pedal.
3	28 (M)	Multiple Schlorsis (Since 2014)	Problem with balance; Feel tired most of the time; pain in arms and shoulders	IP: Average; GM: Expert; ST: Dragon; Apple Siri; AT: Head Tracker, Foot pedal.
4	29 (F)	RSI (Since 2014)	Fatigue; Sore wrists occasionally; Shoulder pain; Pulsing pain in fingers.	GD: Average; IP: Average; ST: Google voice search services; AT: N/A.
5	28 (M)	Tenosinovitis (Since 2016)	Muscle’s weakness; Fatigue; Lack of balance; Unable to use hands	IP: Average; GM: Expert; ST: Dragon software; Google Assistant AT: Tobii eye tracker.
6	36 (F)	RSI (Since 2014)	Muscle’s weakness; Fatigue; pain on wrist and fingers.	IP: Average; GM: Expert; ST: Google speech services; AT: NA.
7	50 (F)	MND (Since 2017)	Problem with balance; Tiredness; unable to move hands.	IP: Average; GM: Average; ST: Google speech services, AT: eye tracker.

8	29 (F)	RSI (Since 2015)	Fatigue; Pinched nerve; Muscle strains; Difficulty in holding stuff.	IP: Expert; GM: Expert; ST: Talon Voice, Google voice search; AT: Eye tracker.
9	34 (M)	MND (Since 2016)	Cannot walk or move hands; lack of balance.	IP: Average; GM: Average; ST: Google Assistant, Samsung Bixby; AT: Eye Tracker, USB Triple Foot Switch Pedal
10	42 (F)	Tendinitis (Since 2019)	Fatigue; Muscle strain; Difficulty when holding stuff.	IP: Average; GM: Expert; ST: Dragon software; Talon; Google Assistant; AT: JellyBean switch
11	22 (M)	RSI (Since 2020)	Throbbing pain effect on hands; wrist pain; shoulder pain.	IP: Average; GM: Average; ST: Apple Siri, Google Assistant; AT: Foot pedal.
12	30 (M)	Spinal Muscular Atrophy (Since 2015)	Uses Powered Chair; unable to move hands and legs; Lack of balance	IP: Average; GM: Average; ST: Dragon software; AT: Eye tracker.

5.4.1.2 Apparatus

The testing sessions were conducted online using Zoom or Microsoft Teams depending on each participant's preference. Participants were informed that they needed to use their own computer and microphone during the testing session. Participants were also given freedom to use whichever device they preferred as a switch input to control the speech recogniser. The study was conducted using the Google Chrome browser (version: 80 or above) to ensure compatibility with the Web Speech API.

5.4.1.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.3: Object Rotation [page 173]). An advertisement was posted on social media networks where those interested were requested to contact the named researcher. A suitable time for the session was then scheduled on participants' preferred online platform for choice (i.e., Teams or Zoom). During the testing session, participants were provided with the web link of the prototype which they were asked to access and then share their screen content with the researcher. The initial pages of prototype consisted of a participant information sheet followed by a consent page which participants were requested to read through and then provide their formal consent by selecting the appropriate confirmation button. Once consent was obtained, participants completed pre-test questions focused on demographic information, details around the nature of their physical impairments, as well as technical experience with interface prototyping, graphical manipulation software and speech technology ((Appendix 8.3: Object Rotation [page 176])).

The same four types of objects used in the elicitation study were utilised again (i.e., images, text, lines, and icons). Participants were initially presented with a training task that required rotating an object (presented in the middle of the canvas) using speech commands such as “left 10” and “right 10”. Once the training task was completed, participants moved onto the main tasks where an object rotated to a specific angle (in the middle of the design canvas) was displayed along with a target placeholder (presented underneath the interaction object) (Figure 5.3). The order of these tasks was counterbalanced to reduce the potential for order bias. There were four main tasks across each of the four different types of graphical assets (i.e., total 16 tasks) where two tasks involved rotating objects across larger distances and two tasks involved rotation at smaller transformations (Table 5.4). The larger transformations again required an interaction object to be rotated at least 180 degrees to reach a target placeholder, while for smaller transformations this distance was 50 degrees or below. The variety of transformations and types of objects were chosen to ensure participants can perform a range of tasks to test the viability of rotating objects via speech. This also presented an opportunity to identify any difference in results using different types of objects and smaller and larger transformation sizes.

After completing a set of four tasks for an object type (e.g., text), participants moved onto the next object type (e.g., image, lines, or icon) and completed the next for tasks. At the end of the testing session, participants were administered the SUS questionnaire and a post-study interview was conducted (Appendix 8.3: Object rotation [Pages 177-178]). Testing sessions were video recorded for further analysis with all testing sessions lasting between 35 to 40 minutes.

Table 5.4 Object types, transformation classifications, starting and final rotation angles

Object Type	Task No.	Transformation Classification	Starting Rotation (degrees) (Interaction Objects)	Final Rotation (degrees) (Target Placeholders)
Image	1	Large	0	180
	2	Large	175	-45
	3	Small	-10	5
	4	Small	-120	-145
Text	1	Large	-60	130
	2	Large	175	-45
	3	Small	15	35
	4	Small	20	-5
Lines	1	Large	10	210
	2	Large	115	-75
	3	Small	25	50
	4	Small	-90	-125
Icons	1	Large	60	275
	2	Large	180	-60
	3	Small	-45	-35
	4	Small	75	35

5.4.1.4 Measures

Task completion time, rotation accuracy, speech recognition performance, and SUS were used to evaluate the object rotation approach developed. Task completion time was measured from when participants clicked the ‘Next Screen’ link in the right sidebar (which started the next task) until they clicked the ‘Next Screen’ link again upon task completion. Rotation accuracies were measured using the differences in the final rotation position of the interaction objects and target placeholders. Speech recognition performance was measured through the total number of speech commands, as well as speech recognition errors which were again categorised into three categories (as highlighted in Section 3.2.3.4 [page 38]):

‘Speech Misrecognition’, ‘System Errors’, and ‘Unsupported commands’. SUS was used to evaluate overall perceptions of usability for the object rotation approach.

5.4.2 Results

No statistical analysis was performed in relation to this exploratory study due to the small sample size.

5.4.2.1 Task Completion Time

The overall mean task completion time across all twelve participants was 11 minutes and 29 seconds (SD= 1.54), while the mean task completion time across all large transformation tasks was 1.85 minutes (SD= 0.63) and 1.11 minutes (SD = 0.49) for small transformations. In terms of image tasks, the mean task completion time across all tasks was 1.75 minutes (SD = 0.61), while for large transformations it was 2.21 minutes (SD = 0.46) and 1.31 minutes (SD = 0.37) for smaller rotation tasks. For text tasks, the mean task completion was 1.43 minutes (SD = 0.68) across all tasks, while for large transformations it was 1.83 minutes (SD = 0.60) and 1.04 minutes (SD = 0.52) for smaller rotation tasks. The mean task completion time across all line tasks was 1.14 minutes (SD = 0.53), while for large transformations it was 1.42 minutes (SD = 0.46) and 0.85 minutes (SD = 0.43) for smaller tasks. In relation to icon tasks, the mean task completion time across all tasks was 1.62 minutes (SD = 0.68), while for large transformations it was 1.96 minutes (SD = 0.68) and 1.28 minutes (SD = 0.49) for smaller tasks.

5.4.2.2 Rotation Accuracy

The overall mean rotation accuracy was 0.83 degrees (SD = 0.85) across all 12 participants, while the mean rotation accuracy across all object types for large transformations was 0.94 degrees (SD = 0.98) and 0.75 degrees (SD = 0.70) for small transformations. The mean rotation accuracy for tasks using images was 0.89 degrees (SD = 0.84), 0.79 degrees (SD = 0.97) for text tasks, 0.81 degrees (SD = 0.83) for line tasks, and 0.85 degrees (SD = 0.76) across all icon tasks. The mean rotation accuracy for large transformation of images tasks was 1.08 degrees (SD = 1.07) and 0.70 degrees (SD = 0.45) for small transformations. The mean rotation accuracy for large transformation of text tasks was 0.91 degrees (SD = 0.99) and 0.70 degrees (SD = 0.97) for small transformations. The mean rotation accuracy for large transformation of line tasks was 0.83 degrees (SD = 0.98) and 0.79 degrees (SD =

0.64) for small transformations. The mean rotation accuracy for large transformation of icon tasks was 0.91 degrees (SD = 0.86) and 0.79 degrees (SD = 0.64) for small transformation tasks.

5.4.2.3 Usability Score

The mean SUS score across all participants was 72.08 (SD = 9.09) which can be labelled as a “Good” level of usability (Bangor et al., 2009).

5.4.2.4 Speech Recognition Performance

Overall, a total of 1164 (SD = 4.42) speech commands were issued across all 12 tasks and participants including 706 (SD = 3.87) commands for large transformations and 458 (SD = 3.29) for small transformations (across all four types of objects). A total of 309 (SD = 5.34) voice commands were issued across all image tasks of which 32 (10.56%) commands were related to ‘Speech Misrecognition’ and 10 (3.23%) to ‘System Errors’. A total of 293 (SD = 4.70) voice commands were issued across all text tasks of which 26 (8.87%) were related to ‘Speech Misrecognition’ and 7 (2.38%) to ‘System Errors’. A total of 257 (SD = 3.04) voice commands were issued across all line tasks of which 22 commands (8.56%) were related to ‘Speech Misrecognition’ and 5 (1.94%) to ‘System Errors’. Finally, a total of 305 (SD = 3.94) commands were issued across all icon tasks of which 29 (9.5%) commands were related to ‘Speech Misrecognition’ and 10 (3.27%) to ‘System Errors’. No utterances related to ‘Unsupported Commands’ were identified across all testing tasks.

5.4.2.5 Qualitative Feedback

Overall, participants provided positive feedback on using speech input for rotating objects on the design canvas. All participants highlighted that the “left/right” commands (in combination with rotation angles such as “left 10” and “right 20”) were simple, easy to use, and short to pronounce:

“The voice commands are basic, straightforward, and easy to understand that how these would work to rotate objects”. (Participant 8)

However, six participants also highlighted that they experienced challenges in estimating the correct rotation angle (especially for larger transformations) and that they had to repeat or issue more commands to position the object to the target location – for example:

“... I liked using simple commands left and right. But it was difficult to get the correct angle value when the distance of object was greater from target place, although I tried to guess the exact rotation angle but couldn't do so and I had to issue more speech commands until object reached at final position". (Participant 3)

"I noticed that when target rotation angle was larger then I had to issue more commands for example, I assumed command "right 120" will take object at exact position but object reached beyond the target position, then I said "left 50" but still could not reach at target then again I estimated more angle values until I finally managed to place object at target". (Participant 10)

No participants mentioned any challenges associated with rotating different object types. Three participants (Participants 5, 7, 11) suggested that some sort of support or visual hints (e.g., guidelines) could reduce the use of speech commands for larger transformations. Three participants (Participants 3, 7, 9) also highlighted that the use of a "flip" command might reduce the number of commands used to rotate objects requiring larger transformation distances:

"… I have used flip option in design applications for rotation tasks and I think if the 'flip' voice command is used then it would help to cover half rotation distance just using this single voice command and so would reduce the distance from target position, then a person would issue few 'left' and 'right' commands to adjust object at target place". (Participant 9)

5.4.3 Summary

Overall, results across the exploratory study highlighted that all participants were able to successfully complete tasks using the speech-controlled object rotation approach. All participants highlighted that they found speech commands "right" and "left" easy to use and effective for rotation tasks, in addition to the interaction approach being rated as exhibiting a good level of usability. The task completion time results suggested that on average participants took longer to rotate objects when larger transformations were required (although no statistical comparison was conducted). This also appears to correlate with the total number of speech commands, where participants used a greater number of commands for tasks required larger transformations.

Moreover, observations during evaluation sessions noted that participants experienced challenges in estimating the correct rotation angle for large transformation tasks and had to issue more commands. In contrast, there were no significant issues identified by participants around estimation of the rotation angle for smaller transformation tasks. There were some issues on occasions related to speech misrecognition (e.g., "left 10" being identified as "let then" and "right 8" as "right ate"), although this was not highlighted as a significant problem by participants. This study therefore validated that the rotation of objects via speech interaction was feasible, although also participants highlighted some clear usability issues (in particular, around estimating transformation sizes). To explore this area further, an additional study was conducted to investigate the efficacy of alternative voice-controlled rotation approaches.

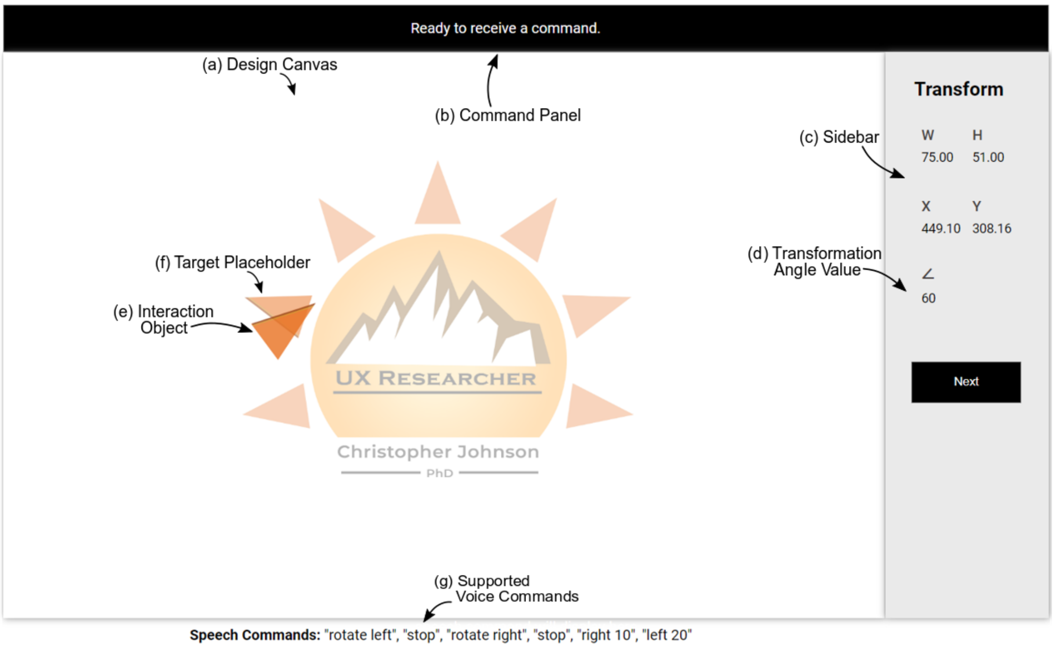
5.5 User Study

The previous exploratory study found that participants were able to successfully complete all rotation tasks using speech interaction. Participants also found speech commands 'short' and 'easy to pronounce', although a key issue around correctly estimating rotation transformation angles was identified. To address this issue, three different voice-controlled rotation approaches were developed and evaluated with participants who have physical impairments.

5.5.1 Research Prototype

The existing research prototype was updated to investigate different object rotation approaches within a design canvas via speech interaction. The application was built using HTML, CSS, and JavaScript – with the Web Speech API (Web Speech API, 2023) again used for recognising speech commands. The prototype interface presented a logo design task for a fictional professional designer (Figure 5.4).

The prototype interface utilised a similar structure and design to those used in Chapters 3 and 4. The interaction object (Figure 5.4 (e)) was presented as an element of a logo design which could be rotated using appropriate voice commands in each condition. A target placeholder (Figure 5.4 (f)) displayed below the interaction object represents the final orientation where the object needs to be placed. Only one active interaction object is rotated on the canvas at a time (for a single task) while all other objects remain deactivated at that time. The supported speech commands (Figure 5.4 (g)) are displayed at the bottom of the canvas to help users in recalling the available speech commands. Switch input (such as a keyboard, mechanical switch, eye tracker, or a foot pedal, etc.) could be used to initiate the speech recogniser before starting each task. Based on the previous exploratory study, the same rotation approach was used again to provide a baseline to compare against two new techniques: 'Fixed-Jumps' and 'Animation'.



Ready to receive a command.

Transform

W 75.00

H 51.00

X 449.10

Y 308.16

∠ 60

Speech Commands: "rotate left", "stop", "rotate right", "stop", "right 10", "left 20"

Figure 5.4 Main object rotation prototype demonstrating interface elements (a) design canvas, (b) command panel, (c) sidebar, (d) transformation angle value, (e) interaction object, (f) target placeholder, and (g) supported voice commands

5.5.1.1 Baseline Rotation

Baseline Rotation uses simple commands such as “right 10” and “left 20” which were also used previously in the exploratory study. When a user issues a command (e.g., “right 10”), the relevant object is rotated 10 degrees in clockwise direction. Similarly, when a voice command such as “left 10” is issued, the interaction object rotates 10 degrees in an anti-clockwise direction. There are no additional commands in this approach to support rotation of objects. Figure 5.5 presents an example of how the objects are rotated using Baseline Rotation.

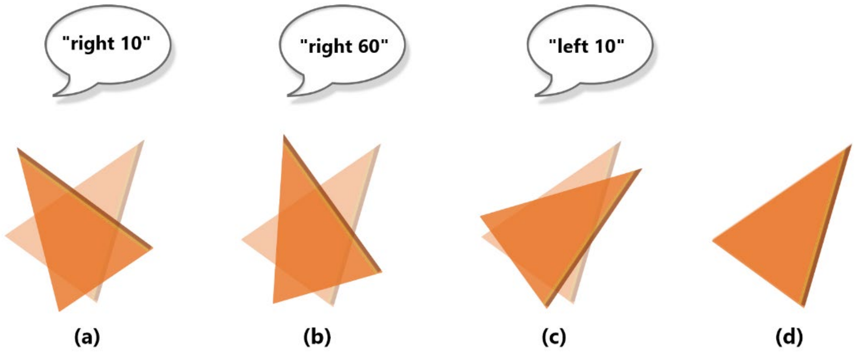


Figure 5.5 Baseline Rotation: (a) the darker object is being rotated in clockwise direction using command “right 10”, (b) user issues command “right 60” to reach towards target rotation position, (c) object goes beyond target position, thus the user issues another “left 10” command, (d) object is successfully positioned at target rotation angle through command

5.5.1.2 Baseline + Fixed-Jumps

This approach continuously rotates the interaction object via 15 degrees ‘jumps’ – to initiate this method, users can issue “rotate left” or “rotate right” commands (Figure 5.6). The object will then rotate in 15-degree increments in the relevant direction until the “stop” command is issued. If the object does not arrive at the intended rotation position after stopping the incremental jumps, the user can then use the Baseline Rotation approach (“left 5” or “right 10”) to adjust the object position over the corresponding target placeholder. This approach is developed around a similar concept used in mainstream creative applications (i.e., Adobe Photoshop, XD, Figma, and Inkscape) where holding the shift key locks the rotation and mouse dragging movements to rotate objects in 15-degree jumps. The motivation behind using this approach was to explore if a continuous rotation approach can potentially reduce the number of speech commands that users need to issue (as identified in the exploratory study), as well as whether this has any impact on perceptions of usability.

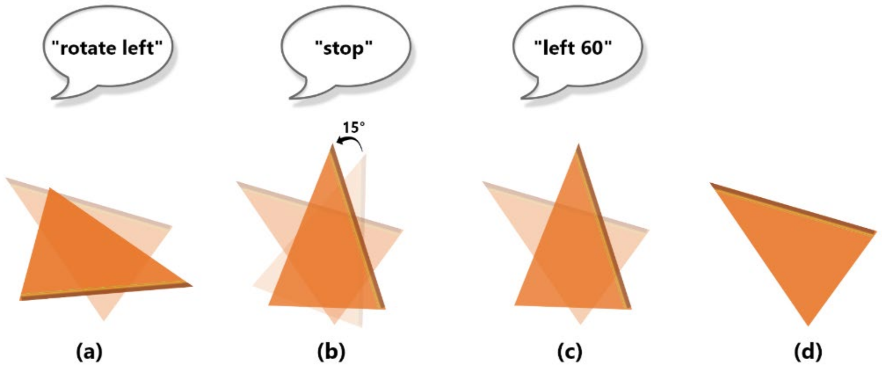


Figure 5.6 Fixed-Jumps approach: (a) user issues commands “rotate left” to rotate object in an anticlockwise direction (b) object is rotated continuously in 15 degrees when the user issues a “stop” as the object reaches closer to the target rotation (c) after stopping the rotation, the user issues a “left 60” command, (d) object is placed at target rotation angle

5.5.1.3 Baseline + Animation

This approach also builds on the Baseline Rotation approach – in particular, when a user issues a command such as “rotate right” or “rotate left”, the object continuously rotates in the relevant direction with smooth motion at a specific speed (without making jumps as in Fixed-Jumps). A user can also issue a “faster” command to rotate an object with increased speed, as well as a “slower” command to reduce the rotation speed. Furthermore, a user can issue a “stop” command to stop the current rotation of an object (Figure 5.7). Users can then issue voice commands associated with Baseline Rotation (e.g., “right 10”, “left 20” etc.) to refine the object position over the target placeholder. A potential benefit of this approach is that it can facilitate users in efficiently rotating objects by reducing the number of speech commands that need to be issued. The option to dynamically alter the rotation speed can also support users in rapidly manipulating objects (especially those that require larger transformations). However, it is also possible that this method may make rotation tasks more tedious if a user is not able to effectively control the transformation speed, thus leading to frustration and usability issues.

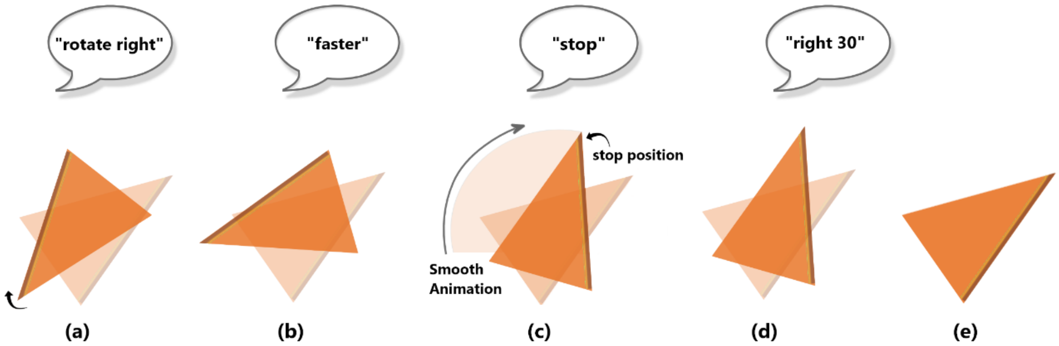


Figure 5.7 Animation Rotation: (a) object is being rotated smoothly (without 15 degrees jumps) in a clockwise direction, (b) user issues "faster" command to increase rotation speed, (c) user stops the rotation close to the target position, (d) user issues command "right 30" to adjust object over target placeholder, (e) object has been adjusted over target rotation position

5.5.2 Methodology

5.5.2.1 Participants

25 participants (15 male and 10 female) with physical impairments were recruited through online advertisements and social media networks (Facebook, LinkedIn, and Slack). Participants were aged between 24 to 50 years (M=33.08, SD=6.97) and were native-English speakers. They were assessed based on their level of experience with graphical design software, prototyping applications and speech interaction technology (Table 5.5).

Table 5.5 Participant information: Physical impairments and condition details; IP = Interface Prototyping (Software); GM = Graphical Manipulation (Software); ST = Speech Technology; AT = Assistive Tools

ID	Age/ Gender	Physical Impairments	Condition Details	Technical Experience
P1	45 (M)	Motor Neurone Disease (MND) Since (2019)	Weakness in Muscles; Lack of balance; uses Powered Chair	IP: Average; GM: Expert; ST: Dragon; Apple Siri; AT: Head Tracker.
P2	24 (F)	Multiple Sclerosis (Since 2020)	Tiredbess; Lack of balance; Tingling sensation in fingers.	IP: Average; GM: Average; ST: Dragon software, Windows speech over; AT: eye tracker.
P3	37 (F)	Repetetitive Strain Injury (RSI) (Since 2014)	Hand tremors; Wrist Pain; Difficulty in using fingers.	IP: Average; GM: Average; ST: Apple Siri, Google voice search; AT: Foot Pedal.
P4	50 (F)	Motor Neurone Disease (MND) (Since 2012)	Muscle’s weakness; Fatigue; Lack of balance; Unable to use hands.	GD: Average; IP: Average; ST: Google Assistant; AT: N/A.
P5	29 (M)	RSI (Since 2017)	Wrist pain; shoulder pain; pain in fingers occassionally.	IP: Expert; GM: Expert; ST: Dragon software; Talon; Google Assistant AT: Trackball mouse.
P6	30 (F)	RSI (Since 2019)	Severe pain in hands when use keyboard; Tiredness in shoulders and upper arms.	IP: Average; GM: Expert; ST: Dragon, Google Assistant; AT: Foot pedal.
P7	26 (M)	RSI (Since 2021)	Tiredness; Numbness in fingers; wrist pain	IP: Expert; GM: Expert; ST: Dragon software, AT: NA.
P8	31 (M)	Tendinitis (Since 2020)	Fatigue; Pinched nerve; Muscle strains; Difficulty in holding stuff.	IP: Average; GM: Expert; ST: Google Assistant; AT: NA.
P9	29 (F)	RSI (Since 2016)	Wrist pain, Pain in shoulders and upper arms; Tiredness; Stiffness in joints.	IP: Average; GM: Average; ST: Google Assistant; AT: NA.
P10	37 (F)	RSI (Since 2010)	Shooting pain in hands and arms; wrist pain; fatigue.	IP: Expert; GM: Average; ST: Dragon software; AT: Foot pedal.
P11	31 (F)	Tenosynovitis (Since 2019)	Joint swelling, wrist pain; pain in fingers; stiffness in hands.	IP: Average; GM: Expert; ST: Dragon software, Apple Siri; AT: Jellybean switch, eye tracker.
P12	39 (M)	MND (Since 2021)	Fatigue in shoulders and arms; lack of balance; uses walking stick.	IP: Average; GM: Expert; ST: Dragon software, Google Assistant; AT: NA.
P13	25 (M)	Spinal Muscular Atrophy (Since 2011)	Cannot walk; Uses powered chair; Unable to move hands and legs; Lack of balance.	IP: Average; GM: Expert; ST: Talon voice, Google Assistant; AT: Eye tracker, Head Pointer.
P14	24 (M)	RSI (Since 2017)	Shooting pain in hands and arms; Tingling; Pain in wrists.	IP: Average; GM: Expert; ST: Google Assistant, Apple Siri; AT: Head Tracker.
P15	43 (M)	Arthritis (Since 2018)	Tiredness; joints pain; weakness in arms and legs; inflammation around joints.	IP: Average; GM: Average; ST: Google Assistant; AT: Eye tracker.
P16	28 (M)	Tendinitis (Since 2018)	Fatigue; Numbness in arms; difficulty when holding stuff.	IP: Expert; GM: Expert; ST: Dragon, Talon Apple Siri; AT: NA.
P17	38 (M)	Motor Neuron Disease (MND) (Since 2017)	Uses walking stick; Arms and shoulders pain; Fatigue.	IP: Average; GM: Expert; ST: Google speech services, Mac voice control; AT: Head Tracker, Eye Tracker.
P18	28 (M)	RSI	Hand tremors; Shooting pain	IP: Average; GM: Expert;

		(Since 2017)	in hands and arms; Pain in wrists; muscle weakness.	ST: Dragon software; AT: NA.
P19	37 (F)	Muliple Sclerosis (Since 2006)	Fatigue; mobility problem; aching body; numbness and tingling in different parts of body	IP: Expert; GM: Expert; ST: Amazon Alexa; Google speech services. AT: eye tracker.
P20	32 (M)	RSI (Since 2018)	Pain in shoulders and arms; Tiredness; numbness in fingers	IP: Average; GM: Expert; ST: Google Home; Apple Siri; AT: NA
P21	30 (M)	RSI (2020)	Weakness in arms, Pain in shoulders, Severe pain when lift arms above head	IP: Expert; GM: Expert; ST: Dragon software, Talon Voice; AT: NA.
P22	35 (F)	RSI (Since 2018)	Pain in forearms and elbows; Throbbing sensation in fingers; joint swelling sometimes.	IP: Expert; GM: Expert; ST: Google Speech Services; AT: NA.
P23	44 (M)	MND (Since 2017)	Difficulty with walking without stick; Fatigue; Lack of balance; Weak grip, Hard to climb stairs.	IP: Average; GM: Expert; ST: Samsung Bixby, Google Assistant; AT: Eye Tracker; Head Tracker.
P24	29 (M)	RSI (Since 2017)	Stiffness of joints; feeling of numbness in fingers; muscles weakness;	IP: Average; GM: Expert; ST: Dragon, Google Search; AT: Foot Pedal.
P25	26 (F)	Muscular Dystrophy (Since 2018)	Lack of balance; frequently falls; muscle pain and stiffness; uses powered chair.	IP: Average; GM: Everage; ST: Dragon; Apple Siri; AT: Eye tracker.

5.5.2.2 Apparatus

Testing sessions were conducted online using video-conferencing platforms such as Zoom, Microsoft Teams, or Skype (depending on participants' preference). Participants were required to use their own computer or laptop with a built-in or external microphone for speech input. Participants were also asked to utilise a switch input (a keyboard or assistive tool of their choice) to enable the speech recogniser. Twelve participants used a physical keyboard (i.e., spacebar key), six used Dragon software via a 'press spacebar' command Dragon, 2023) four utilised a foot pedal, and three used an eye tracker (with an on-screen keyboard). The Google Chrome browser (version 80 or above) was utilised to ensure compatibility with the Web Speech API.

5.5.2.3 Procedure

Ethical approval was obtained from Birmingham City University's ethical committee to conduct this study (Appendix 8.3: Object Rotation [page 173]). The advertisement for the testing session was posted on social media networks where interested users were requested to contact the named researcher. A suitable time for the session was then scheduled on their preferred testing platform. During the evaluation session, participants were provided with a link to the object rotation prototype which they were requested to access on their own device and then share their screen content with the researcher. The initial pages of prototype consisted of a participant information sheet followed by consent page which participants went through and then provided their formal consent by clicking the relevant “confirmation” button. Once consent had been obtained, participants completed a pre-test survey where questions focused on demographic information, graphic design, prototyping and speech technology experience (Appendix 8.3: Object Rotation [Page 176]).

Once the survey was completed, participants were assigned to one of the interaction conditions and started a training task where they were able to freely rotate a single shape (i.e., triangle) within a blank design canvas using the relevant approach (for approximately 5 minutes). Participants then selected a button available in the interface sidebar to move onto the next screen where a screenshot of the first task was displayed to present an overview of what was required before starting the interactive task. Once participants confirmed that they understood the task requirements, they then selected a ‘Start Task’ link and began rotating the interaction object. There were 14 tasks across each interaction condition (i.e., 42 tasks in total) where the order of interaction modes and tasks were counterbalanced to reduce the potential for order bias.

The main task screen consisted of a logo design activity in which each part of the logo had to be rotated. In each task, participants had to rotate a single element of the logo (using the supported speech commands) over the target placeholder beneath the object. The relevant interaction object for each task was initially presented at a different rotation angle than the target placeholder to ensure that participants had to rotate shapes in both clockwise and anticlockwise directions (Figure 5.8). Once participants felt they had accurately completed the rotation activity, they clicked the ‘Next Task’ link within the sidebar to move onto the next task. A screenshot of the next task was then displayed again (prior to starting the task) to ensure participants understood what was required (this process was completed across all tasks). Once all fourteen tasks had been completed for an interaction approach, participants were presented with the SUS form to complete. They then moved onto the next interaction technique and again started with an initial training session, followed by the main tasks, and then completion of the SUS form. After the same process had been completed across all three conditions, a semi-structured interview was held with questions focusing on their perceptions of the different rotation techniques (Appendix 8.3: Object Rotation [page 178]). Testing sessions lasted between 50 to 60 minutes in total and were video recorded for later analysis.

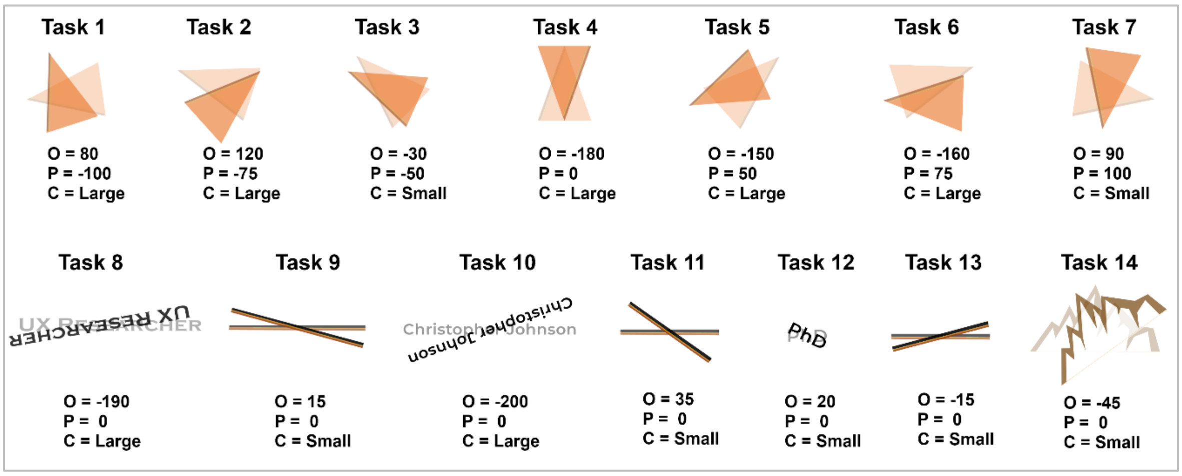


Figure 5.8 Interaction Object (O) – in darker shade, Target Placeholder (P) – semi-transparent object located underneath the interaction object, and rotation transformation category I – large or small

5.5.2.4 Measures

The measures for each interaction approach included task completion time, rotation accuracy, speech recognition performance, and usability. Task completion time was measured in milliseconds and was captured from when a participant selected the 'Start Task' button until they clicked the 'Next Screen' button available in the prototype sidebar. Rotation accuracy was measured through the differences in the final rotation angle of interaction objects and target placeholders. Speech recognition performance was measured through the total number of speech commands issued by participants and speech recognition errors which were again classified into three categories: 'Speech Misrecognition', 'System Error', and 'Unsupported Commands' (as highlighted in Section 3.2.3.4 [page 38]). Moreover, SUS was used to evaluate the perceived usability of each interaction approach.

5.5.3 Results

Shapiro-Wilk's test (Shapiro and Wilk, 1965) for normality (p>0.05) found that task completion time, total number of speech commands, and SUS data were normally distributed, while rotation accuracy and speech recognition error data were not normally distributed. A one-way repeated measures ANOVA and post-hoc paired samples t-tests with Bonferroni correction were utilised to analyse the differences in task completion time, total number of speech commands and SUS score for each interaction approach. A non-parametric Friedman test (Zimmerman and Zumbo, 1993) with Wilcoxon signed rank was used to analyse the differences in rotation accuracy and speech recognition errors between conditions.

5.5.3.1 Task completion time

The mean task completion time for Baseline Rotation was 7.86 minutes (SD=0.62), 8.85 minutes (SD=0.56) for Fixed-Jumps, and 7.52 minutes (SD=0.55) for Animation. A statistically significant difference was found between all three conditions in terms of task completion time (F (2, 48)=171.19, $p<0.001$, partial $\eta^2=0.87$). Post-hoc tests identified a significant difference between Baseline Rotation and Animation ($p<0.05$), Baseline Rotation and Fixed-Jumps ($p<0.05$), as well as Animation and Fixed-Jumps ($p< 0.05$) (Figure 5.9).

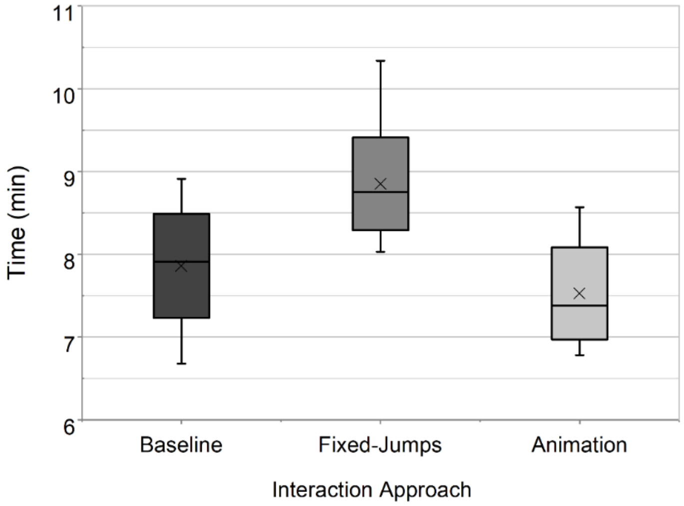


Figure 5.9 Mean Task Completion Time across the three speech-controlled object rotation approaches

5.5.3.2 Rotation Accuracy

The mean rotation accuracy for Baseline Rotation was 0.88 degrees (SD=0.92), 0.91 degrees (SD=0.71) for Fixed-Jumps, and 0.70 degrees (SD=0.73) for Animation. Friedman test results found significant differences in rotation accuracies ($X^2=0.007$, $df=2$, $p<0.05$). The post-hoc Wilcoxon signed rank found significant differences between Baseline Rotation and Animation (Z=-3.38, $p<0.001$), and Animation and Fixed-Jumps (Z=-3.56, $p<0.001$). However, no significant differences were observed between Baseline Rotation and Fixed-Jumps (Z=-0.74, $p=0.45$). Figure 5.10 demonstrates that the mean rotation accuracy of Animation is higher as the difference between the final positions of interaction objects and target placeholders is lower compared to the other conditions.

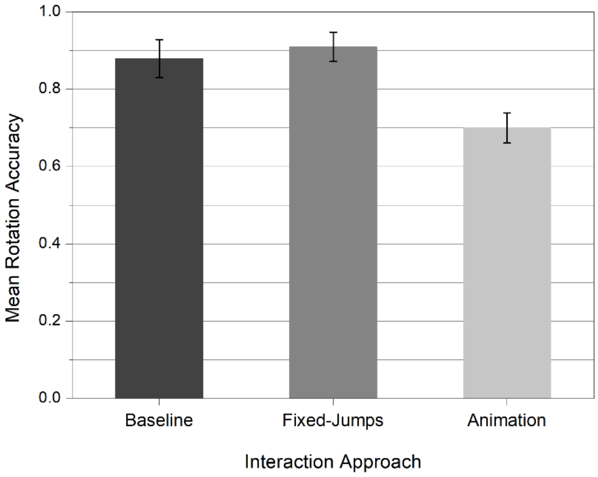


Figure 5.10 The mean rotation accuracy across all three object rotation approaches

5.5.3.3 Speech Recognition Performance

The total number of speech commands issued across all participants for Baseline Rotation was 2704 (SD=12.84), 2968 (SD=10.83) for Fixed-Jumps, and 2553 (SD=15.66) for Animation (Figure 5.11). A statistically significant difference was found between the three conditions in terms of total number of speech commands (F (2, 48)=10.20, $p<0.001$, partial $\eta^2=0.29$). Post-hoc tests highlighted a significant difference between Baseline Rotation and Animation ($p<0.05$), Baseline Rotation and Fixed-Jumps ($p<0.05$), and between Animation and Fixed-Jumps ($p< 0.05$).

There were 181 (6.69%) ‘Speech Misrecognition’ errors for Baseline Rotation, 202 (6.80%) for Fixed-Jumps, and 159 (6.22%) for Animation. Friedman test results found no statistically significant differences for ‘Speech Misrecognition’ across the three interaction approaches ($X^2=0.45$, $df=2$, and $p>0.05$). For ‘System Errors’, there were 68 (2.51%) commands related to Baseline Rotation, 75 (2.53%) associated with Fixed-Jumps and 52 (2.04%) related to Animation. A Friedman test highlighted no statistically significant differences for ‘System Errors’ across the three interaction approaches ($X^2=0.21$, $df=2$, and $p>0.05$). 6 (0.22%) ‘Unsupported Commands’ were issued in Baseline Rotation, 12 (0.37%) in Fixed-Jumps, and 8 (0.34%) in Animation. These included commands such as “left up” instead of “left 30”, as well as the combination of commands such as “rotate right stop” (as opposed to stating “rotate right” and “stop” separately). There were no statistically significant differences across the three conditions for ‘Unsupported Commands’ based on a Friedman test ($X^2=0.32$, $df=2$, and $p>0.05$).

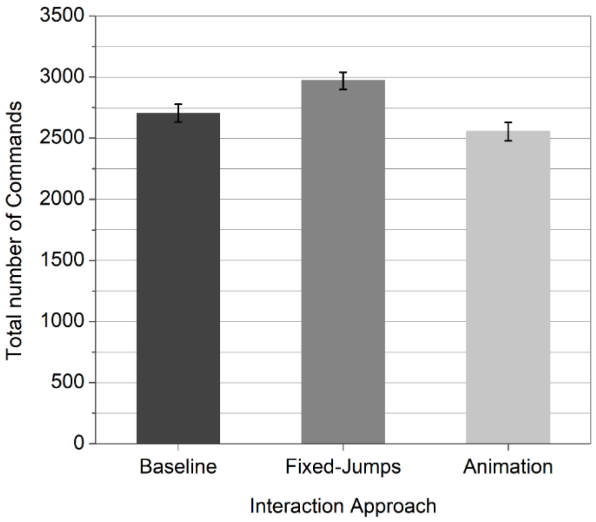


Figure 5.11 Total number of speech commands

5.5.3.4 System Usability Score (SUS)

The mean SUS score for Baseline Rotation was 73.20 (SD=7.72), 70.20 (SD=7.90) for Fixed-Jumps, and 75.20 (SD=8.71) for Animation. The Baseline Rotation and Animation scores can be labelled as "Good" while Fixed-Jumps Rotation can be labelled as "Above Average" (Bangor et al., 2009). A statistically significant difference was found between the three speech interaction approaches in terms of SUS scores (F (2,48)=19.40, $p<0.001$, partial $\eta^2=0.44$). The post-hoc test displayed a significant difference between Baseline Rotation and Animation ($p<0.05$), Baseline Rotation and Fixed-Jumps ($p<0.05$), and between Animation and Fixed-Jumps ($p< 0.05$) (Figure 5.12).

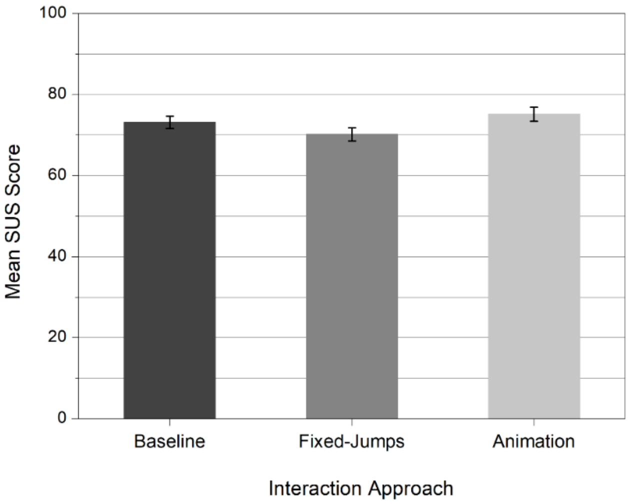


Figure 5.12 The mean SUS score across three rotation interaction approaches

5.5.3.5 Qualitative Feedback

Fifteen participants preferred the Animation approach, eight highlighted a preference for Baseline Rotation, whilst two participants preferred Fixed-Jumps. Participants who provided positive feedback regarding the Animation technique suggested it was both ‘faster’ and more ‘effective’ than Baseline Animation and Fixed-Jumps. Participants also highlighted that the Animation approach enabled them to efficiently rotate objects for larger transformations – for example:

“… liked Animation Rotation because it helps in speeding up the rotation and reduces the effort to reach closer to target position”. (Participant 5)

“I liked Animation Rotation as it provides more control over object by rotating faster and slower based on the situation. When target was at longer rotation position then I used faster command and I was able to quickly reach close to target position”. (Participant 14)

“… I liked that how Animation method evenly rotates object as compared to Fixed-Jumps Rotation where object jumps to a certain angle each time and takes to unintended position”. (Participant 17)

“…reduced the estimation of required rotation angle in comparison to Fixed-Jumps and Baseline but at times it was also tricky to be prepared and assess … when to issue “stop” command”. (Participant 19)

Eight participants who preferred Baseline Rotation commented that they found this approach ‘easy to use’, ‘straightforward’ and having ‘less complexity’ than other approaches – for example:

“…it was easily understandable that how to directly rotate objects in different directions, and I felt relieved as I don’t have to assess when to give ‘stop’ command, so I felt I had more control to decide how much I wanted to rotate object”. (Participant 9)

“… Baseline Rotation was simple and much easier to use because it provides simplistic commands like using only two words left and right with rotation angle as opposed to other two techniques”. (Participant 22)

Whilst two participants preferred Fixed-Jumps, it was generally perceived as cumbersome and difficult to use for object rotation tasks via speech:

“… Fixed-Jumps Rotation was good in a sense that you don’t have to think about exact rotation angle each time as it takes you closer to target position with continuous rotation, but it was also highly likely inaccurate in getting target rotation angle as “stop” command causes delay”. (Participant 6)

Only a single participant highlighted that they felt Fixed-Jumps approach helped them to effectively rotate objects for large transformation tasks:
"I believe Fixed-Jumps was helpful to rotate those objects in which I had to cover a large distance to reach at target because it did not require me to estimate rotation angle each time, I just had to issue stop command and then adjust the object at target as needed". (Participant 13)

5.6 Discussion and Conclusions

This chapter has addressed the research question highlighted in Section 5.2 by developing and evaluating new voice-operated object rotation techniques with people who have physical impairments. An initial elicitation study identified the types of commands that physically impaired users would prefer to use when manipulating the rotation of digital assets within a design context. The findings from this work informed the design of a research prototype that enabled participants to manipulate the orientation of objects using the commands identified. An exploratory study utilising this prototype demonstrated that people with physical impairments could successfully alter the rotation of digital assets, although some key usability challenges were highlighted (e.g., around estimating transformation sizes). To address these challenges, an updated version of the prototype was developed and evaluated with physically impaired users. Results found that the Animation approach was significantly faster, more accurate, and usable that the other two approaches.
Subjective feedback from participants also highlighted a preference for the Animation Rotation technique over Baseline Rotation and Fixed Jumps. However, eight participants still stated a preference for Baseline Rotation due to its simplicity and not having to 'stop' an animation. It appears that a small number of participants felt a certain sense of being 'rushed' when objects were rotating via the Animation approach and had to monitor it closely to choose when to issue the "stop" command (or when it was appropriate to use slower of faster speeds). This coupled with some occasional standard latency issues in terms of the animation stopping after the "stop" command had been issued (which is common in cloud-based services) led to some users preferring Baseline Rotation. However, it is clear overall that there was a preference for Animation in the context of the tasks completed during the study. The study recruited participants with a diverse range of physical impairments and technical experiences although no significant challenges were observed or highlighted by participants around the presented speech-controlled approach.

Whilst Fixed-Jumps still enabled participants to successfully complete all tasks (e.g., positional accuracy was within one degree), it is clear that the majority did not receive this method as positively as the other approaches. A key issue was that the continuous ‘jumps’ did not provide users with the same level of control as the other methods (e.g., the dynamic speed controls in Animation) resulting in users finding this approach a little challenging and tedious to use. It might be that this approach may still provide benefits in other creative design scenarios (e.g., transforming multiple objects simultaneously in a unified way that aligns with jump increments), although further empirical work is required to understand the cases where this technique might better support designers. These findings therefore demonstrate new knowledge around the feasibility of rotating objects using different voice-controlled techniques. Previous studies have explored multimodal approaches where gestures and speech commands (e.g., “spin” and “yaw”) have been used in conjunction (Alibay et al., 2017; Williams et al., 2020), whilst the work presented here utilises a unimodal speech interaction approach. This work is also the first to validate object rotation approaches with people who have physical impairments, thus demonstrating the viability of these methods for this target audience.

In summary, the work detailed in this chapter contributes deeper insights around the viability of voice-controlled rotation approaches to support people with physical impairments with creative work. The next chapter will present the overarching conclusions across the thesis, as well as some of the limitations of the research conducted and interesting areas requiring further investigation in future work.

6 CONCLUSIONS AND FUTURE WORK

The introductory chapter of this thesis highlighted that there has been a lack of work to date around how speech interaction can potentially support people with physical impairments in manipulating and transforming digital assets. The research highlighted in Chapters 3-5 has addressed the limited work conducted in this area to help further develop our understanding of the potential of voice interaction in this domain. This final chapter summarises the main contributions from the research, key limitations, wider implications, and important opportunities for future work.

6.1 Main Contributions

This thesis has focused on the three research questions highlighted in the opening chapter which all focused on exploring the potential of speech interaction to support designers with physical impairments in manipulating digital assets. These questions are highlighted here along with details on how the research conducted has further deepened our knowledge in this field. Table 6.1 and Table 6.2 also present a summary of the studies conducted, along with the key findings.

RQ1. How can voice interaction support the positioning of digital objects around a design canvas?

The research presented in Chapter 3 explored this question via the development and evaluation of three novel object positioning approaches – Speed Control, Location Guides, and Positional Guides. Results highlighted that all three approaches were perceived positively across non-disabled users, although the Location Guides method was perceived as more usable and efficient than the other two approaches. Subjective feedback also highlighted a clear preference for Location Guides with participants highlighting that it was easy and straightforward to use, as well as supported with a combination of larger movements across a design space followed by smaller refinements to facilitate pixel-level positioning of digital assets (i.e., images). The follow-up study with physically impaired users validated that findings were consistent with the first user study as participants were able to use Location Guides to complete all object positioning tasks successfully. There was also consistency in terms of usability ratings given by participants with both groups provided scores that can be labelled as “Excellent” (according to Bangor et al. (2009)). These findings present novel contributions to the field as previous research around object positioning has focused primarily on the use of voice interaction as part of a multimodal approach (Hauptmann, 1989; Hiyoshi and Shimazu, 1994; Elepfandt and Grund, 2012). These studies have utilised only basic positioning commands (e.g., “left”, “right”, and “move object there”) without supporting precise control over digital assets. The multimodal interaction approaches presented in these studies use assistive devices such as gestures, and traditional mouse and keyboard which cause significant challenges for people with physical impairments. Moreover, these studies have also not involved participants with physical impairments to evaluate the developed approaches. The research presented in this thesis (Chapter 3) therefore highlights the feasibility of using voice-only interaction to support people with physical impairments in object positioning on a digital canvas and presents new insights on the strengths and limitations of the novel methods developed.

RQ2. How can voice interaction facilitate the efficient resizing of digital objects?

The research outlined in Chapter 4 addressed this research question primarily via the development and evaluation of three new methods (NoSnap, UserSnap, and AutoSnap) for manipulating the size of digital assets where two of these approaches (UserSnap and AutoSnap) utilised snapping features (tailored for voice interaction) that are widely available in mainstream design applications. A key motivation for investigating snapping techniques was to explore new methods that could overcome some of the challenges identified in an initial exploratory study with non-disabled users (e.g., around estimating transformation sizes when resizing objects). Results across main object resize study with physically impaired users found that AutoSnap was perceived to be more usable, efficient, and accurate in resizing objects when compared with the other two approaches. Feedback also emphasised a strong preference for AutoSnap with participants commenting that it was intuitive, reliable, and reduced object resize duration (in comparison with the other conditions). The findings in relation to this research question present novel contributions to the field through developing the knowledge around the use of speech-only interaction to support object resizing transformation. In particular, whilst early research explored the use of basic resize commands using multimodal interaction (Sedivy and Johnson, 1999;

Williams et al. 2020), no previous work had explored the potential of applying snapping approaches to facilitate the resizing of digital assets (via voice interaction). Object snapping is an essential component of current mainstream creative applications to support precise alignment of resizing of digital objects. Earlier literature around object snapping has utilised traditional inputs such as a mouse to align objects using snap guidelines (Baudisch et al., 2005; Masui, 2001; Ciolfi Felice et al., 2016; Heo et al., 2012), although previous research has not developed and evaluated different voice-controlled snapping techniques to support object resizing. Furthermore, no previous research in relation to object resizing via speech input has focused on evaluating approaches with participants who have physical impairments (Gourdol et al., 1992). The research conducted in this study therefore presents new insights and knowledge around how common snapping approaches utilised in mainstream applications can be tailored for supporting voice-controlled object resizing transformation. These novel approaches can present interaction benefits to support users with physical impairments in efficiently resizing objects on a digital canvas as well as present insights around the strengths and weaknesses of these approaches.

RQ3. How can voice interaction support transformation of digital object orientation?

The studies detailed in Chapter 5 investigated this question further via an initial elicitation study with people who have physical impairments to obtain a sense of the types of commands that might be used to support rotation of digital assets. Whilst a range of speech commands were suggested, results found that participants had a clear preference for commands such as “left/right”. These commands were built into a research prototype and evaluated in an exploratory study with participants who have physical impairments to gain a deeper understanding of the challenges associated with rotation transformations via speech. A key theme to emerge from this study was that participants experienced challenges in accurately estimating the angle of rotation transformations required to place an object in a desired location (thus leading to multiple commands having to be issued to refine the rotation angle). To explore this issue further, three novel voice-controlled rotation approaches were developed (Baseline, Fixed-Jumps, and Animation) and evaluated in a study with 25 participants who have physical impairments. Results found that the Animation condition was more usable, efficient, and accurate when compared with the other approaches. Subjective feedback also illustrated a general preference for this approach with participants highlighting that the ability to dynamically control the speech of rotation presented interaction benefits. Whilst previous studies have explored multimodal interaction where hand gestures and speech commands (i.e., “spin” and “yaw”) were used to rotate objects in a 3D environment (Alibay et al., 2017; Williams et al., 2020), no research to date has explored different unimodal speech interaction techniques for object rotation on a 2D canvas with people who have physical impairments. The speech interaction techniques for object rotation presented in Chapter 5 therefore presents novel contribution and address a gap in the literature around examining the feasibility of object rotation via voice interaction within a 2D creative visual design context.

As a wider perspective, the findings from three research studies exploring object positioning, resizing, and rotation can also have applicability in other related domains. For example, positioning of objects in commercial office software – word processors, presentation applications, operating systems, as well as the potential to support object manipulation on different platforms (e.g., tablets and mobile phones). While the speech interaction approaches presented for object manipulations (positioning, resizing, and rotation) were designed primarily for people with physical impairments, these methods also hold potential to support non-disabled designers. For instance, these approaches can augment a designer’s creative workflow who only rely on using a mouse, keyboard or a stylus for creative activities thus facilitating more efficient transformations (e.g., through enabling users to initiate actions without having to directly select objects and manipulate small transformation handles via a mouse). This can reduce the physical cost of accessing these features, as well as the cognitive load associated with manually locating features (Kim et al., 2019).

Table 6.1 A summary of individual user studies and interaction approaches across each study representing number of tasks and participants

Independent User Studies	Research Studies	Interaction Techniques	No. of Tasks	No. of Participants
Object Positioning	First User Study	Speed Control	9	30 (non-disabled)
		Location Guides	9	
		Positional Guides	9	
	Follow-up Study	Location Guides	9	6 (Disabled)
Object Resize	Exploratory Study	Speech Controlled Resize approach	8	12 (non-disabled)
	Second User Study	NoSnap	10	25 (disabled)
		UserSnap	10	
		AutoSnap	10	
Object Rotation	Elicitation Study	-	12	12 (disabled)
	Exploratory Study	Speech Controlled Rotation approach	12	12 (disabled)
	Third User Study	Baseline	14	25 (disabled)

Table 6.2 Three independent user studies presenting interaction techniques, usability scores, and classification of usability scores according to Bangor et al., (2009)

Independent User Studies	Research Studies	Interaction Techniques	Usability Score	Usability Ratings Bangor et al., (2009)
Object Positioning	First User Study	Speed Control	76.83 (SD = 14.99)	Good
		Location Guides	86.56 (SD = 14.09)	Excellent
		Positional Guides	80.25 (SD = 14.27)	Good
	Follow-up Study	Location Guides	85.42 (SD =12.28)	Excellent
Object Resize	Exploratory Study	Speech Controlled Resize approach	73.12 (SD = 9.42)	Good
	Second User Study	NoSnap	70.40 (SD = 3.72)	Above Average
		UserSnap	73.20 (SD = 11.42)	Good
		AutoSnap	82.00 (SD = 3.75)	Excellent
Object Rotation	Elicitation Study	-	-	-
	Exploratory Study	Speech Controlled Rotation approach	72.08 (SD = 9.09)	Good
	Third User Study	Baseline	73.20 (SD = 7.72)	Good
		Fixed-Jumps	70.20 (SD = 7.90)	Above Average
		Animation	75.20 (SD = 8.71)	Good

6.2 Research Limitations

Whilst this thesis presents multiple novel and original contributions, there are also some limitations that are important to highlight and discuss. For instance, one limitation of the work is the accuracy of speech recognition which is a known challenge within the field and can influence the usability of systems (Alsuraihi and Rigas, 2007; Nishimoto et al., 1995).Whilst the recognition accuracy was high across all studies (approximately 95%), there were still occasions where users had to repeat commands to perform different actions. For instance, words “right”, “snap”, “big”, and “rotate right” were misrecognised and identified as other similar words such as “write”, “nap”, “dig”, and “288 right”. Furthermore, similar sounding alphabetical commands such as “Q” and “U” (which refer to snap point labels in the resize study) were occasionally interpreted incorrectly. Related voice control systems (used by disabled people) such as Talon (Talon, 2023) use a phonetically diverse list of words for typing characters which contain a smaller number of syllables (as compared to NATO phonetic alphabets) (Solanki et al., 2017). It will therefore be important in future work to explore a set of commands that are efficient to pronounce and phonetically diverse (i.e., in terms of containing fewer syllables (Elepfandt and Grund, 2012) being easy to recall, etc.).

A further limitation is that all three studies were conducted over a single testing session. It will be important in future research studies to evaluate these and new techniques in longitudinal studies (Kjærup et al., 2021) where users are exposed to different methods on multiple occasions over extended periods of time. This will provide further insights around how users perceive these approaches and will highlight any additional interaction challenges that may need to be addressed in future (e.g., challenges around vocal fatigue). Moreover, as participants become more familiar with different techniques over time there might be adaptations or optimisations that can be applied to further enhance the efficiency and effectiveness of each transformation approach.

In terms of the testing environment, all studies were conducted via online platforms such as Teams and Zoom which meant that it was not always possible to maintain a fully consistent testing environment. For instance, participants were allowed to use either an external or built-in microphone where the quality of these devices may have differed across users (Modak et al., 2016). Similarly, there was also the possibility of surrounding noise being captured by a microphone which could have potentially affected speech recognition (Pandey et al., 2021; Prodeus and Kukharicheva, 2017). However, despite these environmental limitations, all participants were able to successfully complete tasks across each of the user studies with speech recognition accuracy not being identified as a significant problem during testing sessions. This variety in testing environments also mimics more ‘real-world’ scenarios where users will be working in different locations with varying types of equipment and highlights a degree of robustness in the approaches developed. Another limitation is that studies involved only native English speakers as participants across all user studies. It will therefore be important in future work to further explore the efficacy of the speech interaction approaches in collaboration with participants issuing commands in other languages to evaluate the wider robustness of the system (Laput et al., 2013).

In relation to the first study on object positioning (Chapter 3), a potential limitation of the interaction objects and placeholders used within the study was that they were designed to match the size of images used in common portfolio design tasks (to help create a more realistic design scenario). It will be crucial in future studies to investigate a range of different graphical objects (e.g., shapes, typography, smaller/larger objects, longer/shorter distances

(Creed et al., 2020; Hu et al., 2011; Kamel and Landay, 2000; Sedivy, and Johnson, 1999) and how users find the experience of positioning objects in these different design scenarios. The size of the canvas was also fixed for all tasks whereas it would also be useful moving forward to work with both smaller and larger design spaces to observe whether this has any impact on perceptions of the different positioning approaches. A smaller canvas area would likely require fewer location guides whereas for a larger canvas it might be that further labels are needed (thus potentially impacting on performance and user experience). It is felt that given the positive findings from the research that the approaches would still provide an accessible experience in these different contexts, although this will be important to validate. In terms of the second study on object resizing (Chapter 4), one limitation is the nature of the resizing tasks – in particular, whilst a wide range of common sizes for interface elements were utilised (informed through analysis of visual elements in mainstream applications), it will also be important to cover a wider range of scenarios. For instance, the smallest object size in this study was 35x35 pixels, but it will be useful to explore attempting to resize objects to much smaller sizes to investigate any impact on the efficacy of the approaches developed. Furthermore, it will be important to explore the potential of voice snapping in terms of resizing a wider variety of interface elements (e.g., custom shapes and text) to examine whether this presents any unique interaction challenges that require further consideration. Another area that requires further investigation is the potential for any usability issues that might be presented when the controls of object positioning and resizing are displayed together. For instance, some of the labels might overlap with each other and obscure other interface elements present on the design canvas. Further work is therefore required to explore the simultaneous use of these controls in different scenarios to both understand better any potential issues, as well as opportunities for addressing these challenges.

Finally, for the third study on object rotation (Chapter 5), one limitation is that the task was focused on a specific scenario associated with logo design. It will be important to evaluate the methods developed using different design scenarios and activities (e.g., web design and interface prototyping) to investigate the wider potential of the rotation techniques. Similar to the research presented in Chapters 4 and 5, participants were also required to position objects to align with a pre-defined placeholder. Whilst this provides insights into how participants experienced voice-controlled rotation techniques, further work is also needed around more freeform rotation tasks where users do not have a pre-defined target. It is still anticipated that similar results will be observed in this context, although it could be that new interaction challenges are identified that require further attention.

Another common limitation across the three research studies is that manipulation tasks were designed within a context where all objects were pre-selected. In mainstream creative design applications, objects initially need to be selected before performing other manipulations. Whilst this decision was taken to ensure that object positioning, resizing, and rotation could be explored in detail without potential bias from selection techniques, it will also be important in future work to investigate object selection in combination with other object manipulations in more realistic scenarios.

6.3 Future Work

The speech interaction approaches presented in this thesis hold potential to be integrated into mainstream creative design applications and present opportunities to support the creative workflows of designers with physical impairments, although further research is still required to fully explore the use of speech interaction in creative software. For instance, the positioning approaches presented in Chapter 3 (i.e., Location Guides and Positional Guides) may also present opportunities to support the control of other common features via voice input such as the paint brush and pen/lasso tools (Liu et al., 2018). It will therefore be important in future work to explore the application of these techniques in relation to other design features to examine whether they can support broader control of creative tools. Similarly, whilst the studies presented focused on three key fundamental object manipulations to support creative design work (i.e., positioning, resizing, and rotation), there are also other key interactions associated with creative design work that require further investigation. These include areas such as navigation and selection of objects (Dai et al., 2005; Zhu et al., 2010), simultaneous transformation of multiple objects, and evenly distributing objects to enhance the aesthetic appearance of designs (Xu et al., 2014). These areas have all received little attention to date in relation to voice control within creative domains, so it is essential that further research is conducted around these and other related areas to support the development of more inclusive design applications. The three core studies conducted as part of this thesis focused primarily on the manipulation of images, whereas it will be important in future work to also explore a wider range of different objects such as custom shapes, typography, smaller/larger objects, (Kamel and Landay, 2000; Zhu et al., 2009; Hu et al., 2011; Creed et al., 2020; and Schaadhardt et al., 2021), as well as how users perceive positioning, resizing, and rotating objects in these different design scenarios. Another key interaction issue that requires further investigating is the selection of digital assets on a design canvas (via speech input). This is straightforward when using a mouse, although speech presents some unique issues in that it is unclear how users should refer to an object they wish to select. This can be compounded further when there are potentially hundreds of assets located on a design canvas which could positioned in close proximity to other objects. The use of grids has been explored previously in terms of cursor control (Dai et al., 2003; Zhu et al., 2009) and may present a feasible approach within a creative domain. Another example may be to attach a unique identifier to each item on a canvas (e.g., a number) which users can then state to perform a selection, although this might present challenges around cluttering the interface, as well as how to visualise these 'labels' when multiple objects are located in close proximity.

Similar to the selection of objects, the management of digital assets within a design project is a crucial activity where little research has been conducted around making this inclusive via voice interaction. In mainstream design applications (controlled via mouse and keyboard), objects are normally displayed in a list where each item represents a layer on which the asset is placed. These items can then be manipulated via the list in terms of grouping assets together and placing within other subfolders, moving objects higher or lower in the layout stack (e.g., placing one object/layer on top of another), or to efficiently duplicate specific assets. Whilst these activities are relatively simple to complete via mouse interaction, it is not clear what could be the optimal approaches in terms of managing assets via voice input. Simple fixed commands could be used (e.g., "group", "ungroup", "front", "back", etc.), although these add to the list of commands that users may have to recall thus requiring additional cognitive load (Harada et al., 2009; Schaadhardt et al., 2021; Clark et al., 2019). Natural language commands could help to make this more efficient (e.g., "multi-select line1 and line 2 and group"), although the usability of this approach will likely depend on the quality of the trained system (Cho et al., 2019).

Further work is also required around navigating a project where users might have multiple design canvases or artboards displayed within a single area. In this scenario, controls are required that enable users to efficiently jump to different canvases, panning around the design space, and also zooming in and out of specific areas. Simple navigational commands that were used in the work described in this thesis (i.e., "left", "right, "up", "down", etc.) present a simple solution, but it is likely that a more nuanced solution will be required that enables users to rapidly jump across large spaces to efficiently target a particular design canvas. No work to date has explored how this could be facilitated via voice control and thus represents an important area where further research is required.

A key research challenge in developing inclusive design environments remains around how users can efficiently select and experiment with different colour combinations and palettes. Mainstream design applications typically use a standard interface widget for selecting colours where the full colour space is displayed and can be manipulated via the mouse (Onofre et al., 2018). There has been limited work to date exploring how users can select colours via voice beyond simply stating the name of solid colours (e.g., “red”, “blue”, etc. – Kim et al., 2019; Laput et al., 2013). To produce work to a professional standard, disabled designers need much finer control around the selection of specific colours and the generation of palettes that can be applied consistently across a design project. Further research is therefore required around whether existing approaches can easily be tailored for voice interaction or whether new creative methods are required to better align with this technology.

The ability to create and manipulate creative typography via speech interaction is another area that has received little attention to date. Whilst significant work has explored the use of text creation via voice (Kumar et al., 2012) and multiple commercial applications exist (e.g., Dragon 2023), there has been comparatively less work focused around controlling the visual properties of text (e.g., in terms of font size, font type, letter spacing, line height, etc.). Similarly, whilst there have been studies exploring how to edit text via voice, there has been no work to date that has investigated this in the context of creative design applications. It therefore remains unclear what the optimal approaches are for writing, editing, and manipulating the formatting of text via voice within more creative scenarios. As such, this presents another crucial area where further research is required to support the development of more inclusive design and prototyping applications.

Another key underexplored area is the use of natural language vocal commands to support object positioning. It could be that commands such as “down x pixels”, “place at the top of object x”, or “align horizontally with object x and vertically with object y” would provide users with more flexibility in placing objects around a canvas. The use of natural language has been explored in other related domains (e.g., development work – Paudyal et al. (2020)) which suggests that this may potentially be a fruitful area for further research. It will be particularly important to work with physically impaired designers to understand the scenarios where natural language may have more impact as part of their creative workflow and where it may present barriers. For instance, it might be that the selection of menu items might be best aligned with fixed vocal commands, whereas more creative tasks (e.g., creating, positioning, and transforming multiple objects at one) might be better suited for natural language input. It is also important to note that any interaction benefits in this area may potentially have wider impact and could enhance the creative flow of non-disabled designers who do not currently utilise speech for input purposes.

Alongside further investigation of natural language interaction, the use of non-verbal speech (Harada et al., 2007) in combination with other verbal speech commands may also help to support certain creative tasks. For instance, when users need to make small adjustments to the position or size of objects, the use of continuous vowel sounds (e.g., “aaa”, “eee”) may provide a higher degree of control over excessive repetitive issuing of commands to achieve pixel-level precision (e.g., repeatedly having to state “move left” to position an object into a specific location). However, it will be important to also explore wider issues related to this type of approach such as potential challenges associated with vocal fatigue, as well as from a social perspective to understand the feasibility and acceptance of this type of input method (e.g., users may feel less comfortable issuing vowel sounds around others, as opposed to words and phrases).

Furthermore, the use of alternative multimodal approaches to support creative work is another important area that requires further exploration. This type of approach could help to address some of the known limitations associated with speech interaction (e.g., misrecognition issues – Alsuraihi and Rigas, 2007), thus presenting an enhanced user experience for designers. Previous work has investigated the use of speech in combination with gestures (Neca and Duarte, 2011) and touch input (Laput et al., 2013; Srinivasan et al., 2019), although other modalities may also present benefits. For instance, as highlighted previously, the selection of objects via speech input can present challenges that could be addressed via eye gaze (i.e., fixating gaze on a specific object). However, eye gaze presents its own interaction challenges around the selection of small targets (Jacob and Karn, 2003) and the well-known “Midas Touch” issue (i.e., accidently selecting interface elements where a user’s gaze is fixated – Majaranta and Bulling, 2014). Speech could offer support here by enabling users to confirm that they wish to select an object that they are fixating on (e.g., through issuing a command such as “select”). The combination of different modalities therefore has potential to support interaction in creative domains, although this remains an area where there has been limited work to date. People who have physical impairments that impact speech will be unlikely to benefit from the approaches developed. Other approaches such as the use of short non-verbal commands (e.g., “aaa”, “ooo”) (Harada et al., 2009) may present an alternative for some users, but this will largely be dependent on an individual’s specific requirements and preferences. A combination of eye gaze interaction and non-verbal speech input (Hedeshy et al., 2021) might be a more suitable approach to enhance the user experience, although further work is required to investigate whether this type of approach viable for digital object manipulation to support people with physical impairments.

6.4 Conclusion

This thesis presents new knowledge and contributions to support the development of more inclusive creative visual design experiences for people with physical impairments. In particular, the research conducted demonstrates the potential of voice interaction to make design activities more accessible for people with physical impairments. Further research is now required to investigate how other key design features and tools can be tailored for voice control to help facilitate the production of professional-quality outputs. This is a timely and important field where the wider community needs to be investing more time and resource to ensure disabled people are no longer excluded from utilising mainstream creative design applications.

7 REFERENCES

Aaron, D.L., Patel, A., Kayiaros, S. and Calfee, R. (2011) Four common types of bursitis: diagnosis and management. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, *19*(6), pp. 359-367.

Adamopoulou, E. and Moussiades, L. (2020) Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, p.100006. DOI:https://doi.org/10.1016/j.mlwa.2020.100006

Adobe Photoshop. (2023) Photoshop apps – desktop, mobile, and tablet | Photoshop.com. Available at: https://www.photoshop.com/en [Accessed 26 May 2023]

Adobe. (2023) Designing Out Loud: Announcing Support for macOS Voice Control in Adobe XD. 2023. Available at: https://blog.adobe.com/en/2020/02/11/announcing-mac-os-voice-control-adobe-xd.html?scid=fac788f5-fe6f-4be4-a960-871ac58b5f30&mv=social&mv2=owned_social [Accessed 26 May 2023]

Ali, M., Schur, L. and Blanck, P. (2011) What types of jobs do people with disabilities want?. *Journal of occupational rehabilitation*, *21*(2), pp. 199-210.

Alsuraihi, M.M. and Rigas, D.I. (2007) How effective is it to design by voice?. In: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21, pp. 1-4.

Alusi, S.H., Worthington, J., Glickman, S. and Bain, P.G. (2001) A study of tremor in multiple sclerosis. *Brain*, *124*(4), pp. 720-730.

Ali, A. and Gravino, C. (2019) A systematic literature review of software effort prediction using machine learning methods. *Journal of software: evolution and process*, *31*(10), p-e2211.

Ali, M. and Cardona-Rivera, R.E. (2020) Comparing gamepad and naturally-mapped controller effects on perceived virtual reality experiences. In *ACM Symposium on Applied Perception 2020*, pp. 1-10.

Alibay, F., Kavakli, M., Chardonnet, J. R., & Baig, M. Z. (2017). The usability of speech and/or gestures in multi-modal interface systems. In *Proceedings of the 9th international conference on computer and automation engineering* (pp. 73-77).

Allison, F., Carter, M., Gibbs, M. and Smith, W. (2018) Design patterns for voice interaction in games. In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pp. 5-17.

Ashour, R. and Jankovic, J. (2006) Joint and skeletal deformities in Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. *Movement disorders: official journal of the Movement Disorder Society*, *21*(11), pp. 1856-1863.

Augstein, M., Neumayr, T., Burger, T., Altmann, J., Kurschl, W. and Vrecer, S. (2018) Haptic and Touchless User Input Methods for Simple 3D Interaction Tasks: Interaction Performance and User Experience for People with and Without Impairments. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*, pp. 51-80., Springer, Cham.

Averbuch, A., Bahl, L., Bakis, R., Brown, P., Daggett, G., Das, S., Davies, K., De Gennaro, S., De Souza, P., Epstein, E. and Fraleigh, D. (1987) Experiments with the TANGORA 20,000 word speech recognizer. In: *ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 12, pp. 701-704. IEEE.

Aziz, F., Creed, C., Frutos-Pascual, M. and Williams, I. (2021) Inclusive Voice Interaction Techniques for Creative Object Positioning. In: *Proceedings of the 2021 International Conference on Multimodal Interaction*, pp. 461-469.

Aziz, F., Creed, C., Sarcar, S., Frutos-Pascual, M. and Williams, I. (2022) Voice Snapping: Inclusive Speech Interaction Techniques for Creative Object Manipulation. In *Designing Interactive Systems Conference*, pp. 1486-1496.

Bangor, A., Kortum, P. and Miller, J. (2009) Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), pp. 114-123.

Barker, K.L. (1995) Repetitive Strain Injury: A review of the legal issues. *Physiotherapy*, *81*(2), pp. 103-106.

Baudisch, P., Cutrell, E., Hinckley, K. and Eversole, A. (2005) Snap-and-go: helping users align objects without the modality of traditional snapping. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 301-310.

Becedas, J. (2012) Brain–Machine Interfaces: Basis and Advances. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42(6), pp. 825-836.

Beelders, T.R. and Blignaut, P.J. (2011) The usability of speech and eye gaze as a multimodal interface for a word processor. *Speech Technologies*, pp. 386-404.

Benkerzaz, S., Elmir, Y. and Dennai, A. (2019) A study on automatic speech recognition. *Journal of Information Technology Review*, *10*(3), pp. 80-83.

Benligiray, B., Topal, C., & Akinlar, C. (2019). Slicetype: fast gaze typing with a merging keyboard. *Journal on Multimodal User Interfaces*, *13*, 321-334.

Bennett, C. L., Brady, E., & Branham, S. M. (2018). Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th international acm sigaccess conference on computers and accessibility*, pp. 161-173.

Bier, E.A. and Stone, M.C. (1986) Snap-dragging. *ACM SIGGRAPH Computer Graphics*, *20*(4), pp. 233-240. DOI:https://doi.org/10.1145/15886.15912.

Blain, S., McKeever, P. and Chau, T. (2010) Bedside computer access for an individual with severe and multiple disabilities: a case study. *Disability and Rehabilitation: Assistive Technology*, *5*(5), pp. 359-369.

Brady, E., Morris, M.R., Zhong, Y., White, S. and Bigham, J.P. (2013) Visual challenges in the everyday lives of blind people. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2117-2126.

Burgstahler, S. (2002) Working together: People with disabilities and computer technology. *University of Washington*, DO-IT. Available at: http://easeofaccess.pbworks.com/w/file/fetch/53873128/wtcomp%5B1%5D.pdf [Accessed 26 May 2023].

Brukner, P. (1998) Stress fractures of the upper limb. *Sports Medicine*, *26*(6), pp. 415-424. DOI:https://doi.org/10.2165/00007256-199826060-00004.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), pp. 77-101. DOI:https://doi.org/10.1191/1478088706qp063oa.

Brumberg, J. S., Pitt, K. M., & Burnison, J. D. (2018). A noninvasive brain-computer interface for real-time speech synthesis: The importance of multimodal feedback. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(4), 874-881.

Byrne, J., & Sweetser, P. (2022). Prototyping a Grip Pressure-Sensing Controller for Video Games. In *Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play* (pp. 107-112).

Camargo-Vargas, D., Callejas-Cuervo, M., & Mazzoleni, S. (2021). Brain-computer interfaces systems for upper and lower limb rehabilitation: a systematic review. *Sensors*, *21*(13), 4312.

Cambre, J., Williams, A.C., Razi, A., Bicking, I., Wallin, A., Tsai, J., Kulkarni, C. and Kaye, J. (2021) Firefox voice: An open and extensible voice assistant built upon the web. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-18.

Carr, C.T. and Hayes, R.A. (2015) Social media: Defining, developing, and divining. *Atlantic journal of communication*, *23*(1), pp. 46-65.

Carter, M., Allison, F., Downs, J. and Gibbs, M. (2015) Player identity dissonance and voice interaction in games. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pp. 265-269.

Cave, R. and Bloch, S. (2021) The use of speech recognition technology by people living with amyotrophic lateral sclerosis: a scoping review. *Disability and Rehabilitation: Assistive Technology*, pp. 1-13.

Central London RSI Support Group (2023) Facebook. Available at: https://www.facebook.com/CentralLondonRsiSupportGroup/ [Accessed 26 May 2023]

Çepik, H., Aydın, Ö. And Dalkılıç, G. (2021) Security Vulnerability Assessment of Google Home Connection with an Internet of Things Device. *Multidisciplinary Digital Publishing Institute Proceedings*, *74*(1), p.1.

Chang, M., Huh, M. and Kim, J. (2021) Rubyslippers: Supporting content-based voice navigation for how-to videos. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Pp. 1-14.

Cho, E., Xie, H., Lalor, J.P., Kumar, V. and Campbell, W.M. (2019) Efficient semi-supervised learning for natural language understanding by optimizing diversity. In: *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU).*, pp. 1077-1084, IEEE.

Choi, M., Sakamoto, D. and Ono, T. (2020) Bubble gaze cursor+ bubble gaze lens: Applying area cursor technique to eye-gaze interface. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1-10.

Cincotti, F., Mattia, D., Aloise, F., Bufalari, S., Schalk, G., Oriolo, G., Cherubini, A., Marciani, M.G. and Babiloni, F. (2008) Non-invasive brain–computer interface system: towards its application as assistive technology. *Brain research bulletin*, *75*(6), pp. 796-803.

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J. and R Cowan, B. (2019) The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers*, *31*(4), pp. 349-371.

Ciolfi Felice, M., Maudet, N., Mackay, W.E. and Beaudouin-Lafon, M. (2016) Beyond snapping: Persistent, tweakable alignment and distribution with StickyLines. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 133-144.

Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human-machine communication. *proceedings of the National Academy of Sciences*, *92*(22), 9921-9927.

Corbett, E. and Weber, A. (2016) What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. In: *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*, pp. 72-82.

Creed, C., Beale, R. and Dower, P. (2014) Digital tools for physically impaired visual artists. In: *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pp. 253-254.

Creed, C., Frutos-Pascual, M. and Williams, I., 2020, April. Multimodal Gaze Interaction for Creative Design. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., & Williams, I. (2023). Inclusive AR/VR: accessibility barriers for immersive technologies. *Universal Access in the Information Society*, 1-15.

Dai, L., Goldman, R., Sears, A. and Lozier, J. (2003) Speech-based cursor control: a study of grid-based solutions. *ACM SIGACCESS Accessibility and Computing*, (77-78), pp. 94-101.

Dai, L., Goldman, R., Sears, A. and Lozier, J. (2005) Speech-based cursor control using grids: modelling performance and comparisons with other solutions. *Behaviour & Information Technology*, *24*(3), pp. 219-230.

Dalton, J., Ajayi, V. and Main, R. (2018) Vote goat: Conversational movie recommendation. In *The 41st international acm sigir conference on research & development in information retrieval*, pp. 1285-1288.

Day, C.S., Makhni, E.C., Mejia, E., Lage, D.E. and Rozental, T.D. (2010) Carpal and cubital tunnel syndrome: who gets surgery?. *Clinical Orthopaedics and Related Research®*, *468*(7), pp. 1796-1803.

Dellisanti, M., Fiorentino, M., Monno, G. and Uva, A.E. (2008) Enhanced 3D object snap for CAD modelling on large stereo displays. *International journal of computer applications in technology*, *33*(1), p.54.

Deng, L., Wang, K. and Guido, R.C. (2010) A Semantic and Detection-Based Approach to Speech and Language Processing. *Semantic Computing*, pp. 49-68.

Dalkilic, I. and Kunkel, L.M. (2003) Muscular dystrophies: genes to pathogenesis. *Current opinion in genetics & development*, *13*(3), pp. 231-238.

Dolan, M. J., & Henderson, G. I. (2017). Control devices for electrically powered wheelchairs: prevalence, defining characteristics and user perspectives. *Disability and Rehabilitation: Assistive Technology*, *12*(6), 618-624.

Dove, M.K. (2012) Advancements in assistive technology and AT laws for the disabled. *Delta Kappa Gamma Bulletin*, *78*(4), p.23.

Dragon. (2023) Nuance Communications. *Dragon Speech Recognition – Get More Done by Voice | Nuance*. Available at: https://www.nuance.com/dragon.html [Accessed 26 May 2023]

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods Instruments and Computers*, *34*(4), 455-470.

Duchowski, A.T. (2018) Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, *73*, pp. 59-69. DOI:https://doi.org/10.1016/j.cag.2018.04.002

Elepfandt, M. and Grund, M. (2012) Move it there, or not? The design of voice commands for gaze with speech. In: *Proceedings of the 4th workshop on eye gaze in intelligent human machine interaction.,* pp. 1-3.

Emiliani, P.L. and Stephanidis, C. (2005) Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. *IBM Systems Journal*, *44*(3), pp. 605-619.

Equality act. (2023) Equality Act 2010. Disability. Available at: https://www.legislation.gov.uk/ukpga/2010/15/section/6 [Accessed 26 May 2023]

Evil Controllers. (2023) *Evil Accessible Gaming | Evil Controllers*. [online] Available at: https://www.evilcontrollers.com/accessible-gaming [Accessed 26 May 2023].

Fagarasanu, M., KumarÃ, S., & Narayan, Y. (2005). The training effect on typing on two alternative keyboards. *International Journal of Industrial Ergonomics*, *35*, 509-516.

Feng, J., Karat, C. M., & Sears, A. (2005). How productivity improves in hands-free continuous dictation tasks: lessons learned from a longitudinal study. *Interacting with computers*, *17*(3), 265-289.

Feng, W., Zou, J., Kurauchi, A., Morimoto, C.H. and Betke, M. (2021) Hgaze typing: Head-gesture assisted gaze typing. In: *ACM Symposium on Eye Tracking Research and Applications*, pp. 1-11.

Fernquist, J., Shoemaker, G. and Booth, K.S. (2011) "Oh Snap"–Helping Users Align Digital Objects on Touch Interfaces. In: *IFIP Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg., pp. 338-355.

Figma. (2023) Figma: The collaborative interface design tool. Available at: https://www.figma.com [Accessed 26 May 2023].

Francis, A.L. and Nusbaum, H.C. (1999) Evaluating the quality of synthetic speech. In *Human factors and voice interactive systems*. Springer, Boston, MA., pp. 63-97.

Frisch, M., Kleinau, S., Langner, R. and Dachselt, R. (2011) Grids & guides: multi-touch layout and alignment tools. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1615-1618.

Fuhrer, M.J. (1996) The subjective well-being of people with spinal cord injury: relationships to impairment, disability, and handicap. *Top Spinal Cord Inj Rehabil*, *1*(4), pp. 56-71.

Furini, M., Mirri, S., Montangero, M. and Prandi, C. (2020) Do conversational interfaces kill web accessibility?. In *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*, IEEE., pp. 1-6.

Furui, S. (2005) 50 years of progress in speech and speaker recognition research. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *1*(2), pp. 64-74.

Furui, S. (2010) History and development of speech recognition. In *Speech Technology*. Springer, New York, NY., pp. 1-18.

Girden, E.R. (1992) *ANOVA: Repeated measures* (No. 84). Sage. Available at: https://books.google.co.uk/books?hl=en&lr=&id=JomGKpjnfPcC&oi=fnd&pg=PP7&dq=one+way+repeated+measures+anova&ots=myUtDhTn7B&sig=hIqzBTLl1sivUonLLYkVTnn-Ypg&redir_esc=y#v=onepage&q=one way repeated measures anova&f=false [Accessed 26 May 2023].

Gizatdinova, Y., Špakov, O., Tuisku, O., Turk, M. and Surakka, V. (2018) Gaze and head pointing for hands-free text entry: applicability to ultra-small virtual keyboards. In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pp. 1-9.

Geytenbeek, J.J., Heim, M.M., Vermeulen, R.J. and Oostrom, K.J. (2010) Assessing comprehension of spoken language in nonspeaking children with cerebral palsy: application of a newly developed computer-based instrument. *Augmentative and alternative communication*, *26*(2), pp. 97-107.

Garg, S., Sehgal, A. and Sangwan, S. (2021) Assistive technology is a boon or bane: A case of persons with disabilities. In *Artificial Intelligence and Speech Technology*. CRC Press, pp. 379-385.

Galante, A. and Menezes, P. (2012) A gaze-based interaction system for people with cerebral palsy. *Procedia Technology*, *5*, pp. 895-902.

Goggin, G. (2017) Disability and digital inequalities: Rethinking digital divides with disability theory. In: *Theorizing digital divides*. Routledge, pp. 63-74

Ghannay, S., Caubrière, A., Estève, Y., Camelin, N., Simonnet, E., Laurent, A. and Morin, E. (2018) End-to-end named entity and semantic concept extraction from speech. In: *IEEE Spoken Language Technology Workshop 2018 (SLT)*, pp. 692-699. IEEE.

Gibbon, D., Moore, R. and Winski, R. eds. (1997) *Handbook of standards and resources for spoken language systems*. Walter de Gruyter.

Gibbons, C.J., Thornton, E.W. and Young, C.A. (2013) The patient experience of fatigue in motor neurone disease. *Frontiers in psychology*, *4*, p. 788.

Göbl, B., Kriglstein, S. and Hlavacs, H. (2021) Conversational Interfaces in Serious Games: Identifying Potentials and Future Research Directions based on a Systematic Literature Review. *CSEDU (1)*, pp. 108-115.

Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In *the Mind's Eye*. North-Holland. Amsterdam, pp. 493-516.

Giunchi, D., Sztrajman, A., James, S. and Steed, A. (2021) Mixing Modalities of 3D Sketching and Speech for Interactive Model Retrieval in Virtual Reality. In *ACM International Conference on Interactive Media Experiences*, pp. 144-155.

Gonçalves, D., Rodrigues, A., Richardson, M. L., de Sousa, A. A., Proulx, M. J., & Guerreiro, T. (2021). Exploring asymmetric roles in mixed-ability gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Gourdol, A.P., Nigay, L., Salber, D. and Coutaz, J. (1992) Two Case Studies of Software Architecture for Multimodal Interactive Systems: VoicePaint and a Voice-enabled Graphical Notebook. *Engineering for Human-Computer Interaction*, *92*, pp. 271-84. Available at: https://dl.acm.org/doi/10.5555/647103.717584

Grauman, K., Betke, M., Lombardi, J., Gips, J., & Bradski, G. R. (2003). Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, *2*, 359-373.

Grewal, H. S., Matthews, A., Tea, R., Contractor, V., & George, K. (2018). Sip-and-puff autonomous wheelchair for individuals with severe disabilities. In *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 705-710). IEEE.

Gu, X., Cao, Z., Jolfaei, A., Xu, P., Wu, D., Jung, T. P., & Lin, C. T. (2021). EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM transactions on computational biology and bioinformatics*, *18*(5), 1645-1666.

Guccione, A.A. (1994) Arthritis and the process of disablement. *Physical Therapy*, *74*(5), pp. 408-414.

Gyory, P. (2022). Creating Platforms to Support Craft and Creativity in Game Controller Design. In *Creativity and Cognition* (pp. 708-710).

Gygli, M., & Ferrari, V. (2020). Efficient object annotation via speaking and pointing. *International Journal of Computer Vision*, *128*(5), 1061-1075.

Hackbarth, K. R. (2017). *A Modular Architecture for Control and Communication Assistive Technology Devices* (Doctoral dissertation, Bowling Green State University).

Halsey, M. (2015) Introducing Accessibility in Windows 10. In: *The Windows 10 Accessibility Handbook*. Apress, Berkeley, CA., pp. 1-15.

Hansen, P.J., Hansen, D.W., and Johansen A.S. (2001). Bringing Gaze-based Interaction Back to Basics. In Proceedings of Universal Access in Human-Computer

Interaction (UAHCI 2001), New Orleans, Louisiana. Available at: https://www.researchgate.net/publication/221100181_Bringing_gaze-based_interaction_back_to_basics [Accessed 26 May 2023]

Harada, S., Saponas, T.S. and Landay, J.A. (2007). VoicePen: Augmenting pen input with simultaneous non-linguisitic vocalization. In: *Proceedings of the 9th international conference on Multimodal interfaces*, pp. 178-185. DOI:https://doi.org/10.1145/1322192.1322225

Harada, S., Wobbrock, J.O. and Landay, J.A. (2007) Voicedraw: a hands-free voice-driven drawing application for people with motor impairments. In: *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pp. 27-34. DOI:https://doi.org/10.1145/1296843.1296850

Harada, S., Wobbrock, J.O., Malkin, J., Bilmes, J.A. and Landay, J.A. (2009) Longitudinal study of people learning to use continuous voice-based cursor control. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 347-356.

Harada, S., Wobbrock, J.O. and Landay, J.A. (2011) Voice games: investigation into the use of non-speech voice input for making computer games more accessible. In: *IFIP Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg, pp. 11-29.

Harada, S., Landay, J.A., Malkin, J., Li, X. and Bilmes, J.A. (2006) The vocal joystick: evaluation of voice-based cursor control techniques. In: *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 197-204.

Harada, S., Landay, J.A., Malkin, J., Li, X. and Bilmes, J.A. (2008) The Vocal Joystick: evaluation of voice-based cursor control techniques for assistive technology. *Disability and Rehabilitation: Assistive Technology*, *3*(1-2), pp. 22-34.

Harris, M., Cox, K.C., Musgrove, C.F. and Ernstberger, K.W. (2016) Consumer preferences for banking technologies by age groups. *International Journal of Bank Marketing*.

Hassan, M., Shimizu, Y., Hada, Y. and Suzuki, K. (2022) Joy-Pros: A Gaming Prosthesis to Enable Para-Esports for Persons with Upper Limb Deficiencies. *IEEE Access*, *10*, pp. 18933-18943.

Hauptmann, A.G. (1989) Speech and gestures for graphic image manipulation. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 241-24.

Hiyoshi, M. and Shimazu, H. (1994) Drawing pictures with natural language and direct manipulation. In *COLING 1994 Volume 2: The 15th International Conference on Computational Linguistics*. DOI:https://doi.org/10.3115/991250.991262

Heck, L., Hakkani-Tür, D., Chinthakunta, M., Tur, G., Iyer, R., Parthasacarthy, P., Stifelman, L., Shriberg, E. and Fidler, A. (2013) Multimodal conversational search and browse. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2013/08/slam_cameraReady2.pdf [Accessed 26 May 2023]

Hedeshy, R., Kumar, C., Menges, R., & Staab, S. (2021). Hummer: Text entry by gaze and hum. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-11).

Heikkilä, H. (2013) EyeSketch: A drawing application for gaze control. In: *Proceedings of the 2013 Conference on Eye Tracking South Africa*, pp. 71-74.

Heo, S., Lee, Y.K., Yeom, J. and Lee, G. (2012) Design of a shape dependent snapping algorithm. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 2207-2212.

Hertzum, M., & Hornbæk, K. (2010). How age affects pointing with mouse and touchpad: A comparison of young, adult, and elderly users. *Intl. Journal of Human–Computer Interaction*, *26*(7), 703-734.

Honeycutt, L. (2003) Researching the use of voice recognition writing software. *Computers and Composition*, *20*(1), pp. 77-95.

House, B., Malkin, J. and Bilmes, J. (2009) The VoiceBot: a voice controlled robot arm. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Pp. 183-192.

Hu, R., Zhu, S., Feng, J. and Sears, A. (2011) Use of speech technology in real life environment. In: *International Conference on Universal Access in Human-Computer Interaction*. Springer, Berlin, Heidelberg, pp. 62-71.

Huckauf, A. and Urbina, M.H. (2011) Object selection in gaze controlled systems: What you don't look at is what you get. *ACM Transactions on Applied Perception (TAP)*, *8*(2), pp. 1-14.

Hyrskykari, A., Istance, H. and Vickers, S. (2012) Gaze gestures or dwell-based interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 229-232.

Inkscape. (2023) Draw Freely | Inkscape Design Tool. Available at: https://inkscape.org/ [Accessed 26 May 2023]

Igarashi, T. and Hughes, J.F. (2001) Voice as sound: using non-verbal voice input for interactive control. In: *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pp. 155-156.

Ismail, N. A., & O'Brien, E. A. (2008). Enabling multimodal interaction in web-based personal digital photo browsing. In *2008 International Conference on Computer and Communication Engineering* (pp. 907-910). IEEE.

Jacob, R.J. and Karn, K.S. (2003) Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: *The mind's eye*, pp. 573-605. DOI:https://doi.org/10.1016/B978-044451020-4/50031-1

Jang, M., Choi, J., & Lee, S. (2010). A customized mouse for people with physical disabilities. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility* (pp. 281-282).

Johnson, M., Lapkin, S., Long, V., Sanchez, P., Suominen, H., Basilakis, J. and Dawson, L. (2014) A systematic review of speech recognition technology in health care. *BMC medical informatics and decision making*, *14*(1), pp. 1-14.

Kamel, H.M. and Landay, J.A. (2000) A study of blind drawing practice: creating graphical information without the visual channel. In: *Proceedings of the fourth international ACM conference on Assistive technologies*, pp. 34-41.

Kambeyanda, D., Singer, L., & Cronk, S. (1997). Potential problems associated with use of speech recognition products. *Assistive Technology*, *9*(2), 95-101.

Kane, S. K., Morris, M. R., Perkins, A. Z., Wigdor, D., Ladner, R. E., & Wobbrock, J. O. (2011). Access overlays: improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 273-282).

Kalagi, S., Machado, J., Carvalho, V., Soares, F. and Matos, D. (2017) Brain computer interface systems using non-invasive electroencephalogram signal: A literature review. In: *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. *IEEE,* pp. 1578-1583.

Kasprowski, P. (2022). Eye Tracking Hardware: Past to Present, and Beyond. In *Eye Tracking: Background, Methods, and Applications* (pp. 31-48). New York, NY: Springer US.

Kim, J. W., & Ritter, F. E. (2015). Learning, forgetting, and relearning for keystroke-and mouse-driven tasks: Relearning is important. *Human–Computer Interaction*, *30*(1), 1-33.

Kim, Y.S., Dontcheva, M., Adar, E. and Hullman, J. (2019) Vocal shortcuts for creative experts. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-14. DOI:https://doi.org/10.1145/3290605.3300562

Komal Aziz, F. (2021) Inclusive Visual Design to Support People with Physical Impairments. In: *Companion Publication of the 2021 ACM Designing Interactive Systems Conference*, pp. 20-22.

Kong, F., Sahadat, M. N., Ghovanloo, M., & Durgin, G. D. (2019). A stand-alone intraoral tongue-controlled computer interface for people with tetraplegia. *IEEE transactions on biomedical circuits and systems*, *13*(5), 848-857.

Kumar, A., Paek, T. and Lee, B. (2012) Voice typing: a new speech interaction model for dictation on touchscreen devices. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2277-2286.

Kwan, C. and Betke, M. (2011) Camera Canvas: Image editing software for people with disabilities. In: *International Conference on Universal Access in Human-Computer Interaction*. Springer, Berlin, Heidelberg, pp. 146-154.

Kim, M., Kim, B.H. and Jo, S. (2014) Quantitative evaluation of a low-cost noninvasive hybrid interface based on EEG and eye movement. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(2), pp. 159-168.

Karimullah, A.S. and Sears, A. (2002) Speech-based cursor control. In: *Proceedings of the fifth international ACM conference on Assistive technologies*, pp. 178-185.

Klamka, K., Siegel, A., Vogt, S., Göbel, F., Stellmach, S., & Dachselt, R. (2015). Look & pedal: Hands-free navigation in zoomable information spaces through gaze-supported foot input. In *Proceedings of the 2015 ACM on international conference on multimodal interaction* (pp. 123-130).

Klopfenstein, L.C., Delpriori, S., Malatini, S. and Bogliolo, A. (2017) The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In: *Proceedings of the 2017 conference on designing interactive systems*, pp. 555-565.

Koester, H., Simpson, R. and Mankowski, J. (2013) Software wizards to adjust keyboard and mouse settings for people with physical impairments. *The journal of spinal cord medicine*, *36*(4), pp. 300-312.

Keates, S. and Trewin, S. (2005) Effect of age and Parkinson's disease on cursor positioning using a mouse. In: *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pp. 68-75.

Koch, K. (2017) Stay in the box! Embedded assistive technology improves access for students with disabilities. *Education Sciences*, *7*(4), p.82.

Kjærup, M., Skov, M.B., Nielsen, P.A., Kjeldskov, J., Gerken, J. and Reiterer, H. (2021) Longitudinal Studies in HCI Research: A Review of CHI Publications From 1982–2019. *Advances in Longitudinal HCI Research*, pp. 11-39.

Kumar, Y. and Singh, N. (2019) A comprehensive view of automatic speech recognition system-a systematic literature review. In: *2019 international conference on automation, computational and technology management (ICACTM)*. IEEE, pp. 168-173.

Koester, M.C., George, M.S. and Kuhn, J.E. (2005) Shoulder impingement syndrome. *The American journal of medicine*, *118*(5), pp. 452-455.

Kolb, S.J. and Kissel, J.T. (2011) Spinal muscular atrophy: a timely review. *Archives of neurology*, *68*(8), pp. 979-984.

Laviola, J.J. and Katzourin, M. (2007) An exploration of non-isomorphic 3D rotation in surround screen virtual environments. In: *IEEE symposium on 3D user interfaces 2007*. IEEE.

Lee, S. H., Lee, Y. E., & Lee, S. W. (2022). Toward imagined speech based smart communication system: potential applications on metaverse conditions. In *2022 10th International Winter Conference on Brain-Computer Interface (BCI)* (pp. 1-4). IEEE.

Leonardi, M., Bickenbach, J., Ustun, T.B., Kostanjsek, N. and Chatterji, S. (2006) The definition of disability: what is in a name?. *The Lancet*, *368*(9543), pp. 1219-1221.

Liu, Z., Thompson, J., Wilson, A., Dontcheva, M., Delorey, J., Grigg, S., Kerr, B. and Stasko, J. (2018) Data Illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Laput, G.P., Dontcheva, M., Wilensky, G., Chang, W., Agarwala, A., Linder, J. and Adar, E. (2013) Pixeltone: A multimodal interface for image editing. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2185-2194. DOI:https://doi.org/10.1145/2470654.2481301

Law, M., Steinwender, S. and Leclair, L. (1998) Occupation, health and well-being. *Canadian Journal of Occupational Therapy*, *65*(2), pp. 81-91.

Leijten, M. and Van Waes, L. (2005) Writing with speech recognition: The adaptation process of professional writers with and without dictating experience. *Interacting with Computers*, *17*(6), pp. 736-772.

López, G., Quesada, L. and Guerrero, L.A. (2017) Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In *International conference on applied human factors and ergonomics*. Springer, Cham., pp. 241-250.

Lazarou, I., Nikolopoulos, S., Petrantonakis, P.C., Kompatsiaris, I. and Tsolaki, M. (2018) EEG-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st Century. *Frontiers in human neuroscience*, *12*, p. 14.

Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q. and Martinez, A. (2019) Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, *51*(4), pp. 984-997.

Lugano, G. (2017) Virtual assistants and self-driving cars. In: *15th International Conference on ITS Telecommunications 2017 (ITST)*. IEEE, pp. 1-5.

Lutteroth, C., Penkar, M. and Weber, G. (2015) Gaze vs. mouse: A fast and accurate gaze-only click alternative. In: *Proceedings of the 28th annual ACM symposium on user interface software & technology*, pp. 385-394.

Macik, M., Cerny, T., & Slavik, P. (2014). Context-sensitive, cross-platform user interface generation. *Journal on Multimodal User Interfaces*, *8*, 217-229.

Magee, J. J., Scott, M. R., Waber, B. N., & Betke, M. (2004). Eyekeys: A real-time vision interface based on gaze detection from a low-grade video camera. In *2004 Conference on Computer Vision and Pattern Recognition Workshop* (pp. 159-159). IEEE.

Maggiorini, D., Granato, M., Ripamonti, L. A., Marras, M., & Gadia, D. (2019). Evolution of game controllers: Toward the support of gamers with physical disabilities. In *Computer-Human Interaction Research and Applications: First International Conference, CHIRA 2017, Funchal, Madeira, Portugal, October 31–November 2, 2017, Revised Selected Papers 1* (pp. 66-89). Springer International Publishing.

Martin, J. (2020). The Practical Applications of Video Games Beyond Entertainment. Available at: https://commons.clarku.edu/cgi/viewcontent.cgi?article=1062&context=sps_masters_papers [Accessed 14 April 2023]

Mihara, Y., Shibayama, E. and Takahashi, S. (2005) The migratory cursor: accurate speech-based cursor movement by moving multiple ghost cursors using non-verbal vocalizations. In: *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 76-83.

Masui, T. (2001) HyperSnapping. In: Proceedings IEEE Symposia on Human-Centric Computing Languages and Environments (Cat. No. 01TH8587) IEEE, pp. 188-194.

Majaranta, P. and Räihä, K.J. (2002) Twenty years of eye typing: systems and design issues. In: *Proceedings of the 2002 symposium on Eye tracking research & applications,* pp. 15-22.

Majaranta, P., Ahola, U. K., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 357-360).

Majaranta, P. ed. (2011) Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: Advances in Assistive Technologies. IGI Global. DOI:https://doi.org/10.4018/978-1-61350-098-9

Macdonald, S.J. and Clayton, J. (2013) Back to the future, disability and the digital divide. *Disability & Society*, *28*(5), pp. 702-718.

Mattar, A.A., Hitzig, S.L. and McGillivray, C.F. (2015) A qualitative study on the use of personal information technology by persons with spinal cord injury. *Disability and Rehabilitation*, *37*(15), pp. 1362-1371.

Matter, R., Harniss, M., Oderud, T., Borg, J. and Eide, A.H. (2017) Assistive technology in resource-limited environments: a scoping review. *Disability and Rehabilitation: Assistive Technology*, *12*(2), pp. 105-114.

Meng, J., Zhang, J. and Zhao, H. (2012) Overview of the speech recognition technology. In: *fourth international conference on computational and information sciences 2012*. IEEE, pp. 199-202.

Microsoft Teams (2023) Microsoft Teams | Group Chat, Team Chat & Collaboration. *Microsoft Teams*. Available at: https://www.microsoft.com/en-gb/microsoft-teams/group-chat-software [Accessed 26 May 2023]

Mohapatra, S. (2021) Human and computer interaction in information system design for managing business. *Information Systems and e-Business Management*, *19*(1), pp. 1-11.

Modak, S., Vikmani, S., Shah, S. and Kurup, L. (2016) Voice driven dynamic generation of webpages. In: *International Conference on Computing Communication Control and automation 2016 (ICCUBEA)*. IEEE, pp. 1-4

Miller, R.G., Mitchell, J.D. and Moore, D.H. (2012) Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane database of systematic reviews*, (3).

Majaranta, P. and Bulling, A. (2014) Eye tracking and eye-based human–computer interaction. In: *Advances in physiological computing*. Springer, London, pp. 39-65.

Mauri, C., Granollers, T., Lorés, J. and García, M. (2006) Computer vision interaction for people with severe movement restrictions. *Human Technology*, *2*(1), pp. 38-54.

Mougharbel, I., El-Hajj, R., Ghamlouch, H. and Monacelli, E. (2013) Comparative study on different adaptation approaches concerning a sip and puff controller for a powered wheelchair. In *2013 Science and Information Conference*. IEEE, pp. 597-603.

Müller, P., Buschek, D., Huang, M.X. and Bulling, A. (2019) Reducing calibration drift in mobile eye trackers by exploiting mobile phone usage. In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1-9.

Munteanu, C., Baecker, R., Penn, G., Toms, E. and James, D. (2006) The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 493-502.

Nayyar, A., Dwivedi, U., Ahuja, K., Rajput, N., Nagar, S. and Dey, K. (2017) OptiDwell: intelligent adjustment of dwell click time. In *Proceedings of the 22nd international conference on intelligent user interfaces*, pp. 193-204.

Neca, J. and Duarte, C. (2011) Evaluation of gestural interaction with and without voice commands. In: *Proceedings of IHCI 2011-IADIS Conference on Interfaces and Human Computer Interaction*. Available at: https://www.researchgate.net/publication/256007049 [Accessed 26 May 2023]

Necas, M. (1996) Musculoskeletal symptomnatology and repetitive strain injuries in diagnostic medical sonographers. *J Diagn Med Sonogr*, *12*(6), pp. 266-273.

Nishimoto, T., Shida, N., Koayashi, T. and Shirai, K. (1995) Improving human interface drawing tool using speech, mouse and keyboard. In: *Proceedings 4th IEEE International Workshop on Robot and Human Communication*. IEEE, pp. 107-112. DOI:https://doi.org/10.1109/roman.1995.531944

O'Neil, B.A., Forsythe, M.E. and Stanish, W.D. (2001) Chronic occupational repetitive strain injury. *Canadian Family Physician*, *47*(2), pp. 311-316.

One-Handed DualSense. (2023) *The Controller Project*. [online] Available at: https://thecontrollerproject.com/mods/one-handed-dualsense/ [Accessed 13 April 2023].

Onofre, E. and Van Der Meulen, V.I.N.C.E.N.T., (2018) Color Me Surprised: An Interactive Color Picker for UI/UX Designers. Available at: http://downloads.vincentmvdm.com/colormesurprised.pdf [Accessed 26 May 2023]

Oshita, M., & Ishikawa, H. (2012). Gamepad vs. touchscreen: a comparison of action selection interfaces in computer games. In *Proceedings of the Workshop at SIGGRAPH Asia* (pp. 27-31).

Paudyal, B., Creed, C., Frutos-Pascual, M. and Williams, I. (2020) Voiceye: A multimodal inclusive development environment. In: *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pp. 21-33.

Pan, L., Yu, C., He, Z., & Shi, Y. (2023). A Human-Computer Collaborative Editing Tool for Conceptual Diagrams. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-29).

Pandey, L., Hasan, K. and Arif, A.S. (2021) Acceptability of Speech and Silent Speech Input Methods in Private and Public. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Pape, T.L.B., Kim, J. and Weiner, B. (2002) The shaping of individual meanings assigned to assistive technology: a review of personal factors. *Disability and rehabilitation*, *24*(1-3), pp. 5-20.

Poupyrev, I., Weghorst, S. and Fels, S. (2000) Non-isomorphic 3D rotational techniques. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 540-547.

Pausch, R. and Leatherby, J.H. (1991) An empirical study: Adding voice input to a graphical editor. In *Journal of the American voice input/output society*. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.4829 [Accessed 26 May 2023]

Pousada, T., Pareira, J., Groba, B., Nieto, L. and Pazos, A. (2014) Assessing mouse alternatives to access to computer: a case study of a user with cerebral palsy. *Assistive Technology*, *26*(1), pp. 33-44.

Pradhan, A., Mehta, K. and Findlater, L. (2018) "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In: *Proceedings of the 2018 CHI Conference on human factors in computing systems*, pp. 1-13.

Prodeus, A. and Kukharicheva, K. (2017) Automatic speech recognition performance for training on noised speech. In: *2nd International Conference on Advanced Information and Communication Technologies 2017 (AICT)*. IEEE, pp. 71-74.

Purington, A., Taft, J.G., Sannon, S., Bazarova, N.N. and Taylor, S.H. (2017) “Alexa is my new BFF” Social Roles, User Satisfaction, and Personification of the Amazon Echo. In: *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, pp. 2853-2859.

Pienimäki, T.T., Siira, P.T. and Vanharanta, H. (2002) Chronic medial and lateral epicondylitis: a comparison of pain, disability, and function. *Archives of physical medicine and rehabilitation*, *83*(3), pp. 317-321.

Radwan, A., Kallasy, T., Monroe, A., Chrisman, E., & Carpenter, O. (2018). Benefits of alternative computer mouse designs: A systematic review of controlled trials. *Cogent Engineering*, *5*(1), 1521503.

Rai, S.N., Birla, H., Singh, S.S., Zahra, W., Rathore, A.S., Dilnashin, H. and Singh, S.P. (2019) Pathophysiology of the Disease Causing Physical Disability. In: *Biomedical Engineering and its Applications in Healthcare*. Springer, Singapore, pp. 573-595.

Raisamo, R. and Räihä, K.J. (1996) A new direct manipulation technique for aligning objects in drawing programs. In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pp. 157-164.

Raja, D.S. (2016) Bridging the disability divide through digital technologies. *Background paper for the World Development report*.

Ramadan, Z., F Farah, M. and El Essrawi, L. (2021) From Amazon. Com to Amazon. Love: How Alexa is redefining companionship and interdependence for people with special needs. *Psychology & Marketing*, *38*(4), pp. 596-609.

Ramolete, G. I. L., Almirante, J. I. F., Mondragon, J. M., Ting, C. A. R., Cohen, M. A. P. C., & Custodio, B. P. (2020). Physical design assessment of the nintendo switch controller configurations. In *Advances in Ergonomics in Design: Proceedings of the AHFE 2020 Virtual Conference on Ergonomics in Design, July 16–20, 2020, USA* (pp. 198-205). Springer International Publishing.

Rao, J., Ture, F. and Lin, J. (2018) What do viewers say to their TVs? An analysis of voice queries to entertainment systems. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1213-1216.

Reddy, D.R. (1976) Speech recognition by machine: A review. *Proceedings of the IEEE*, *64*(4), pp. 501-531.

Reeves, B., Ram, N., Robinson, T.N., Cummings, J.J., Giles, C.L., Pan, J., Chiatti, A., Cho, M.J., Roehrick, K., Yang, X. and Gagneja, A. (2021) Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human–Computer Interaction*, *36*(2), pp. 150-201.

Reichle, J. (2011) Evaluating assistive technology in the education of persons with severe disabilities. *Journal of Behavioral Education*, *20*(1), pp. 77-85.

Reitmaier, T., Wallington, E., Kalarikalayil Raju, D., Klejch, O., Pearson, J., Jones, M., Bell, P. and Robinson, S. (2022) Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers. In: *CHI Conference on Human Factors in Computing Systems*, pp. 1-17.

Rempel, D., Nathan-Roberts, D., Chen, B. Y., & Odell, D. (2009). The effects of split keyboard geometry on upper body postures. *Ergonomics*, *52*(1), 104-111.

Röddiger, T., Clarke, C., Wolffram, D., Budde, M., & Beigl, M. (2021). EarRumble: Discreet Hands-and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Röddiger, T., King, T., Roodt, D. R., Clarke, C., & Beigl, M. (2022). OpenEarable: Open Hardware Earable Sensing Platform. In *Proceedings of the 1st International Workshop on Earable Computing, EarComp* (Vol. 22, pp. 29-34).

Ronzhin, A. and Karpov, A. (2005) Assistive multimodal system based on speech recognition and head tracking. In: *2005 13th European Signal Processing Conference*. IEEE, pp. 1-4

Rotariu, C., Costin, H., Bozomitu, R.G., Petroiu-Andruseac, G., Ursache, T.I. and Cojocaru, C.D. (2019) New assistive technology for communicating with disabled people based on gaze interaction. In: *E-Health and Bioengineering Conference (EHB) 2019,* IEEE. Pp. 1-4

Rebman Jr, C.M., Aiken, M.W. and Cegielski, C.G. (2003) Speech recognition in the human–computer interface. *Information & Management*, *40*(6), pp. 509-519.

Robinson, E.L., Park, G., Shalini, S., Levins, T., Lane, K.R., Skubic, M. and Markway, B. (2019) Harnessing voice-assisted technology and in-home sensors to manage older adult health: A user preference study. *Innovation in Aging*, *3*(Supplement_1), pp. S926-S927.

Rough, D., Vertanen, K., & Kristensson, P. O. (2014). An evaluation of Dasher with a high-performance language model as a gaze communication method. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (pp. 169-176).

Saktheeswaran, A., Srinivasan, A. and Stasko, J. (2020) Touch? Speech? Or touch and speech? Investigating multimodal interaction for visual network exploration and analysis. *IEEE transactions on visualization and computer graphics*, *26*(6), pp. 2168-2179.

Schaadhardt, A., Hiniker, A. and Wobbrock, J.O. (2021) Understanding Blind Screen-Reader Users' Experiences of Digital Artboards. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-19.

Scherer, M.J. (2002) The change in emphasis from people to person: introduction to the special issue on Assistive Technology. *Disability and rehabilitation*, *24*(1-3), pp. 1-4.

Scherer, R., Siddiq, F. and Tondeur, J. (2019) The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, pp. 13-35.

Schmidbauer, O., Casacuberta, F., Castro, M.J., Hegerl, G., Höge, H., Sanchez, J.A. and Zlokarnik, I. (1993) Articulatory representation and speech technology. *Language and Speech*, *36*(2-3), pp. 331-351.

Scope. (2023) *Disability facts and figures | Disability charity Scope UK*. [online] Available at: https://www.scope.org.uk/media/disability-facts-figures/ [Accessed 26 May 2023].

Schenk, S., Dreiser, M., Rigoll, G. and Dorr, M. (2017) GazeEverywhere: enabling gaze-only user interaction on an unmodified desktop PC in everyday scenarios. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3034-3044.

Schultz, T., Wand, M., Hueber, T., Krusienski, D. J., Herff, C., & Brumberg, J. S. (2017). Biosignal-based spoken communication: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(12), 2257-2271.

Sears, A., Lin, M. and Karimullah, A.S. (2002) Speech-based cursor control: understanding the effects of target size, cursor speed, and command selection. *Universal Access in the Information Society*, *2*(1), pp. 30-43.

Sears, A., Feng, J., Oseitutu, K. and Karat, C.M. (2003) Hands-free, speech-based navigation during dictation: difficulties, consequences, and solutions. *Human-computer interaction*, *18*(3), pp. 229-257.

Sears, A., Young, M. and Feng, J. (2009) Physical disabilities and computing technologies: an analysis of impairments. In: *Human-Computer Interaction*. CRC Press, pp. 87-110.

Sedivy, J. and Johnson, H. (1999) Supporting creative work tasks: the potential of multimodal tools to support sketching. In: *Proceedings of the 3rd conference on Creativity & cognition*, pp. 42-49.

Sengupta, K., Ke, M., Menges, R., Kumar, C. and Staab, S. (2018) Hands-free web browsing: enriching the user experience with gaze and voice modality. In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Application,* pp. 1-3.

Sereshkeh, A. R., Yousefi, R., Wong, A. T., & Chau, T. (2018). Online classification of imagined speech using functional near-infrared spectroscopy signals. *Journal of neural engineering*, *16*(1), 016005.

Shahsavari, H., Matourypour, P., Ghiyasvandian, S., Ghorbani, A., Bakhshi, F., Mahmoudi, M. and Golestannejad, M. (2020) Upper limb amputation; Care needs for reintegration to life: An integrative review. *International journal of orthopaedic and trauma nursing*, *38*, p.100773.

Shapiro, S.S. and Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), pp. 591-611. DOI:https://doi.org/10.2307/2333709

Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-computer interfaces in medicine. In *Mayo clinic proceedings* (Vol. 87, No. 3, pp. 268-279). Elsevier.

Silsbee, P.L. and Bovik, A.C. (1996) Computer lipreading for improved accuracy in automatic speech recognition. *IEEE Transactions on Speech and Audio Processing*, *4*(5), pp. 337-351.

Sindhwani, S., Lutteroth, C. and Weber, G. (2019) ReType: Quick text editing with keyboard and gaze. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Singh, H., & Singh, J. (2018). Real-time eye blink and wink detection for object selection in HCI systems. *Journal on Multimodal User Interfaces*, *12*, 55-65.

Skovsgaard, H., Mateo, J.C., Flach, J.M. and Hansen, J.P. (2010) Small-target selection with gaze alone. In *Proceedings of the 2010 symposium on eye-tracking research & applications*, pp. 145-148.

Silva, S., Almeida, N., Pereira, C., Martins, A. I., Rosa, A. F., Oliveira e Silva, M., & Teixeira, A. (2015). Design and development of multimodal applications: a vision on key issues and methods. In *Universal Access in Human-Computer Interaction. Access to Today's Technologies: 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part I 9* (pp. 109-120). Springer International Publishing.

Slobodenyuk, N. (2016). Towards cognitively grounded gaze-controlled interfaces. *Personal and Ubiquitous Computing*, *20*, 1035-1047.

Solanki, S., Krishnan, G., Sampath, V. and Polakis, J. (2017) In (cyber) space bots can hear you speak: Breaking audio captchas using ots speech recognition. In: *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, pp. 69-80.

Song, Y., Morency, L.P. and Davis, R. (2012) Multimodal human behavior analysis: learning correlation and interaction across modalities. In: *Proceedings of the 14th ACM international conference on Multimodal interaction*, pp. 27-30.

Sonix. (2023) *A brief history of speech recognition | Sonix*. [online] Available at: https://sonix.ai/history-of-speech-recognition [Accessed 26 May 2023].

Sotolongo, N. and Copulsky, J. (2018) Conversational marketing: Creating compelling customer connections. *Applied Marketing Analytics*, *4*(1), pp. 6-21.

Sporka, A.J., Kurniawan, S.H., Mahmud, M. and Slavík, P. (2006) Non-speech input and speech recognition for real-time control of computer games. In: *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 213-220.

Sporka, A.J., Kurniawan, S.H. and Slavik, P. (2006) Acoustic control of mouse pointer. *Universal Access in the Information Society*, *4*(3), pp. 237-245.

Srinivasan, A., Dontcheva, M., Adar, E. and Walker, S. (2019) Discovering natural language commands in multimodal interfaces. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 661-672.

Sugano, Y. and Bulling, A. (2015) Self-calibrating head-mounted eye trackers using egocentric visual saliency. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 363-372.

Szczechowicz, J. and Mazurek, A. (2018) Keyboards types of selected electronic devices and the occurrence of pain in the shoulder girdle and upper limb–a review of literature. Available at: https://web.archive.org/web/20201124170106id_/https://rehmed.pl/api/files/view/600076.pdf [Accessed 26 May 2023]

Talon. (2023) Talonvoice. Talon. Available at: https://talonvoice.com. [Accessed 26 May 2023]

Tanwear, A., Liang, X., Paz, E., Böhnert, T., Ghannam, R., Ferreira, R., & Heidari, H. (2022). Spintronic eyeblink gesture sensor with wearable interface system. *IEEE Transactions on Biomedical Circuits and Systems*, *16*(5), 779-792.

Timmer, B.H., Hickson, L. and Launer, S. (2015) Adults with mild hearing impairment: Are we meeting the challenge?. *International journal of audiology*, *54*(11), pp. 786-795.

Tiric-Campara, M., Krupic, F., Biscevic, M., Spahic, E., Maglajlija, K., Masic, Z., Zunic, L. and Masic, I. (2014) Occupational overuse syndrome (technological diseases): carpal tunnel syndrome, a mouse shoulder, cervical pain syndrome. *Acta Informatica Medica*, *22*(5), p. 333.

Titz, J., Scholz, A., & Sedlmeier, P. (2018). Comparing eye trackers by correlating their eye-metric data. *Behavior Research Methods*, *50*, 1853-1863.

Trewin, S., Keates, S. and Moffatt, K. (2006) Developing steady clicks: a method of cursor assistance for people with motor impairments. In: *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 26-33.

Trewin, S., Swart, C. and Pettick, D. (2013) Physical accessibility of touchscreen smartphones. In: *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1-8.

Turk, M. (2014) Multimodal interaction: A review. *Pattern recognition letters*, *36*, pp. 189-195. DOI: https://doi.org/10.1016/j.patrec.2013.07.003

Udayashankar, A., Kowshik, A. R., Chandramouli, S., & Prashanth, H. S. (2012). Assistance for the paralyzed using eye blink detection. In *2012 Fourth International Conference on Digital Home* (pp. 104-108). IEEE.

Van der Kamp, J. and Sundstedt, V. (2011) Gaze and voice controlled drawing. In: *Proceedings of the 1st conference on novel gaze-controlled applications*, pp. 1-8. DOI:https://doi.org/10.1145/1983302.1983311

Vassilev, V., Phipps, A., Lane, M., Mohamed, K. and Naciscionis, A. (2020) Two-factor authentication for voice assistance in digital banking using public cloud services. In: *10th International Conference on Cloud Computing, Data Science & Engineering 2020 (Confluence)* (pp. 404-409). IEEE.

Vessoyan, K., Smart, E., Steckle, G., & McKillop, M. (2023). A Scoping Review of Eye Tracking Technology for Communication: Current Progress and Next Steps. *Current Developmental Disorders Reports*, *10*(1), 20-39.

Vitrikas, K., Dalton, H. and Breish, D. (2020) Cerebral palsy: an overview. *American family physician*, *101*(4), pp. 213-220.

Web Speech API (2023). *MND | Developer.mozilla.org – Web APIs* [online] Available at: <https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API> [Accessed 26 May 2023].

Weinstein, A. M. (2010). Computer and video game addiction—a comparison between game users and non-game users. *The American journal of drug and alcohol abuse*, *36*(5), 268-276.

Wibirama, S., Murnani, S. and Setiawan, N.A. (2020) Spontaneous gaze gesture interaction in the presence of noises and various types of eye movements. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1-5.

Wilcox, M.J., Norman-Murch, T., Oberstein, J.S., Volkmann, M.A., Wagner, D.K., Musselwhite, C.R., Malena, E. and Weber, C.A. (1999) Assistive Technology: Tips, Tools, and Techniques. A Parent Resource Manual.

Wobbrock, J. and Myers, B. (2006) Trackball text entry for people with motor impairments. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 479-488.

Wobbrock, J.O., Morris, M.R. and Wilson, A.D. (2009) User-defined gestures for surface computing. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1083-1092.

Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, *3*(3), pp. 1-27.

Xbox. (2023) *Xbox Adaptive Controller | Xbox*. [online] Available at: https://www.xbox.com/en-GB/accessories/controllers/xbox-adaptive-controller [Accessed 26 May 2023].

Xu, P., Fu, H., Igarashi, T. and Tai, C.L. (2014) Global beautification of layouts with interactive ambiguity resolution. In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 243-252.

Wang, L., Zhang, Y. and Feng, J. (2005) On the Euclidean distance of images. *IEEE transactions on pattern analysis and machine intelligence*, *27*(8), pp. 1334-1339. DOI:https://doi.org/10.1109/TPAMI.2005.165

Wentzel, J., Junuzovic, S., Devine, J., Porter, J., & Mott, M. (2022). Understanding How People with Limited Mobility Use Multi-Modal Input. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-17).

WHO. (2023) World Health Organization. *Disability and health*. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/disability-and-health [Accessed 26 May 2023].

Williams, L.J. and Abdi, H. (2010) Fisher’s least significant difference (LSD) test. *Encyclopedia of research design*, *218*(4), pp.840-853. Available at: https://www.researchgate.net/publication/242181775 [Accessed 26 May 2023]

Williams, A.S., Garcia, J. and Ortega, F. (2022) Errata to “Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation”. *IEEE Transactions on Visualization and Computer Graphics*, *28*(7), pp. 2808-2808.

Williams, A.S. and Ortega, F.R. (2020) Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. *Proceedings of the ACM on Human-Computer Interaction*, *4*(ISS), pp. 1-21.

Williams, A.S., Garcia, J. and Ortega, F. (2020) Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation. *IEEE Transactions on Visualization and Computer Graphics*, *26*(12), pp. 3479-3489.

Wirz, M., Müller, R. and Bastiaenen, C. (2010) Falls in persons with spinal cord injury: validity and reliability of the Berg Balance Scale. *Neurorehabilitation and neural repair*, *24*(1), pp. 70-77.

Wobbrock, J. O., Aung, H. H., Myers, B. A., & Lopresti, E. F. (2005). Integrated text entry from power wheelchairs. *Behaviour & information technology*, *24*(3), 187-203.

Woods, V., Hastings, S., Buckle, P., & Haslam, R. (2003). Development of non-keyboard input device checklists through assessments. *Applied ergonomics*, *34*(6), 511-519.

Woolson, R.F. (2007) Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pp. 1-3.

Young, V. and Mihailidis, A. (2010) Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology*, *22*(2), pp. 99-112.

Yuan, H., Li, Y., Yang, J., Li, H., Yang, Q., Guo, C., ... & Shu, X. (2021). State of the Art of Non-Invasive Electrode Materials for Brain–Computer Interface. *Micromachines*, *12*(12), 1521.

Zargham, N., Pfau, J., Schnackenberg, T. and Malaka, R. (2022) “I Didn’t Catch That, But I’ll Try My Best”: Anticipatory Error Handling in a Voice Controlled Game. In *CHI Conference on Human Factors in Computing Systems*, pp. 1-13.

Zhang, X., Lu, C., Yin, J., Xie, H., & Liu, T. (2020). The Study of Two Novel Speech-Based Selection Techniques in Voice-User Interfaces. *IEEE Access*, *8*, 217024-217032.

Zhu, S., Ma, Y., Feng, J. and Sears, A. (2009) Speech-Based Navigation: Improving Grid-Based Solutions. In *IFIP Conference on Human-Computer Interaction*. Springer, Berlin, Heidelberg., pp. 50-62.

Zhu, S., Feng, J. and Sears, A. (2010) Investigating grid-based navigation: The impact of physical disability. *ACM Transactions on Accessible Computing (TACCESS)*, *3*(1), pp. 1-30.

Zielasko, D., Freitag, S., Rausch, D., Law, Y. C., Weyers, B., & Kuhlen, T. W. (2015). Blowclick: A non-verbal vocal input metaphor for clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (pp. 20-23).

8 APPENDICES

APPENDIX 8.1: OBJECT POSITIONING (CHAPTER 3)

Ethical Approval:

BIRMINGHAM CITY
University

Faculty of Computing, Engineering & the Built Environment Research Office
Millennium Point, Curzon Street
Birmingham
B4 7XG

BCU_ethics@bcu.ac.uk

26/Aug/2020

Miss Farkhandah Aziz Komal

farkhandah.komal@mail.bcu.ac.uk

Dear Farkhandah Aziz ,

Re: Komal /7539 /R(B) /2020 /Aug /CEBE FAEC - Speech-based object manipulation

Thank you for your application and documentation regarding the above activity. I am pleased to take Chair's Action and approve this activity.

Provided that you are granted Permission of Access by relevant parties (meeting requirements as laid out by them), you may begin your activity.

I can also confirm that any person participating in the project is covered under the University's insurance arrangements.

Please note that ethics approval only covers your activity as it has been detailed in your ethics application. If you wish to make any changes to the activity, then you must submit an Amendment application for approval of the proposed changes.

Examples of changes include (but are not limited to) adding a new study site, a new method of participant recruitment, adding a new method of data collection and/or change of Project Lead.

Please also note that the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee should be notified of any serious adverse effects arising as a result of this activity.

If for any reason the Committee feels that the activity is no longer ethically sound, it reserves the right to withdraw its approval. In the unlikely event of issues arising which would lead to this, you will be consulted.

Keep a copy of this letter along with the corresponding application for your records as evidence of approval.

If you have any queries, please contact BCU_ethics@bcu.ac.uk

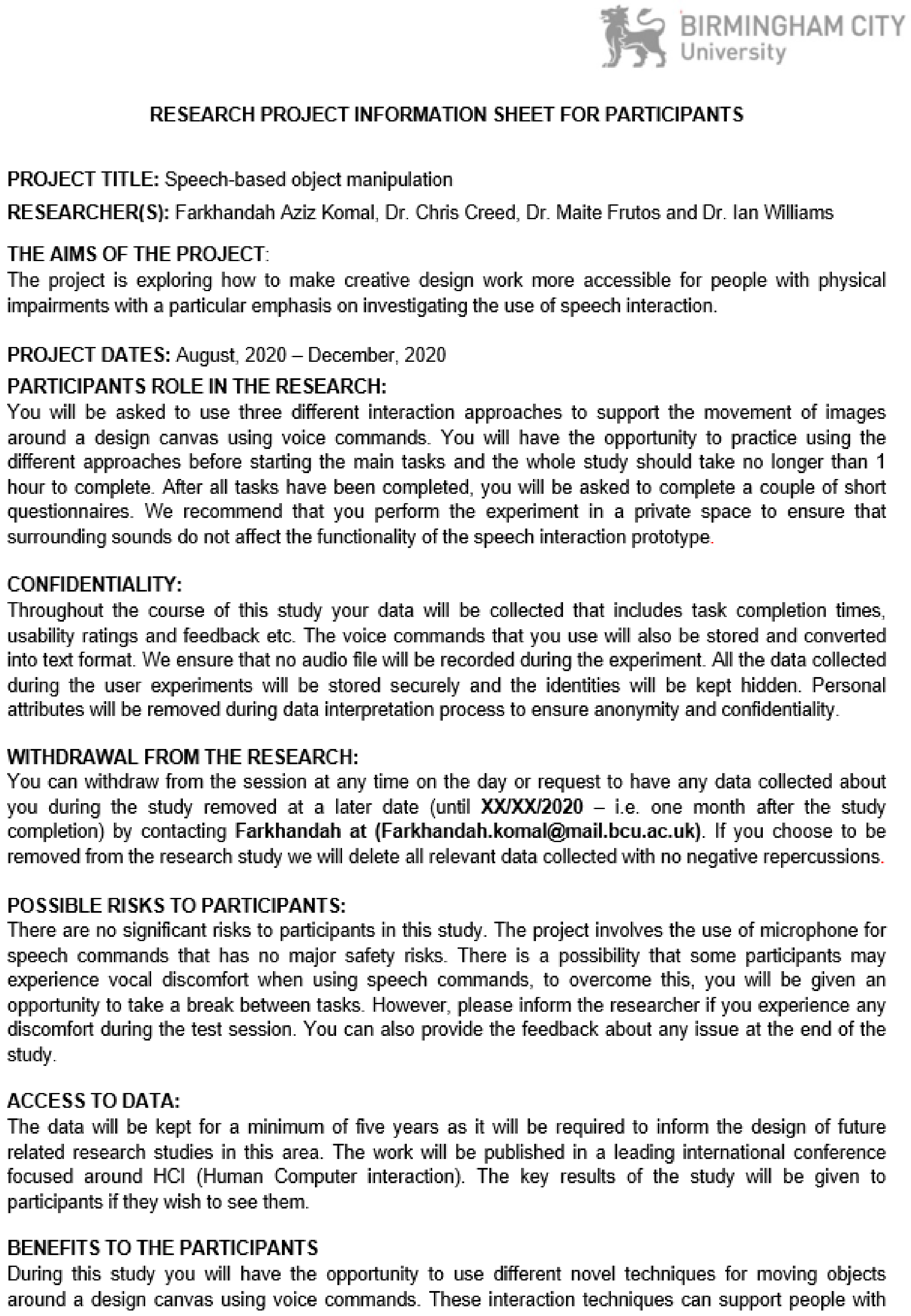
I wish you every success with your activity.

Yours Sincerely,

Professor Peter Larkham

On behalf of the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee

Participants Information Sheet:



BIRMINGHAM CITY
University

RESEARCH PROJECT INFORMATION SHEET FOR PARTICIPANTS

PROJECT TITLE: Speech-based object manipulation

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos and Dr. Ian Williams

THE AIMS OF THE PROJECT:
The project is exploring how to make creative design work more accessible for people with physical impairments with a particular emphasis on investigating the use of speech interaction.

PROJECT DATES: August, 2020 – December, 2020

PARTICIPANTS ROLE IN THE RESEARCH:
You will be asked to use three different interaction approaches to support the movement of images around a design canvas using voice commands. You will have the opportunity to practice using the different approaches before starting the main tasks and the whole study should take no longer than 1 hour to complete. After all tasks have been completed, you will be asked to complete a couple of short questionnaires. We recommend that you perform the experiment in a private space to ensure that surrounding sounds do not affect the functionality of the speech interaction prototype.

CONFIDENTIALITY:
Throughout the course of this study your data will be collected that includes task completion times, usability ratings and feedback etc. The voice commands that you use will also be stored and converted into text format. We ensure that no audio file will be recorded during the experiment. All the data collected during the user experiments will be stored securely and the identities will be kept hidden. Personal attributes will be removed during data interpretation process to ensure anonymity and confidentiality.

WITHDRAWAL FROM THE RESEARCH:
You can withdraw from the session at any time on the day or request to have any data collected about you during the study removed at a later date (until **XX/XX/2020** – i.e. one month after the study completion) by contacting **Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk)**. If you choose to be removed from the research study we will delete all relevant data collected with no negative repercussions.

POSSIBLE RISKS TO PARTICIPANTS:
There are no significant risks to participants in this study. The project involves the use of microphone for speech commands that has no major safety risks. There is a possibility that some participants may experience vocal discomfort when using speech commands, to overcome this, you will be given an opportunity to take a break between tasks. However, please inform the researcher if you experience any discomfort during the test session. You can also provide the feedback about any issue at the end of the study.

ACCESS TO DATA:
The data will be kept for a minimum of five years as it will be required to inform the design of future related research studies in this area. The work will be published in a leading international conference focused around HCI (Human Computer interaction). The key results of the study will be given to participants if they wish to see them.

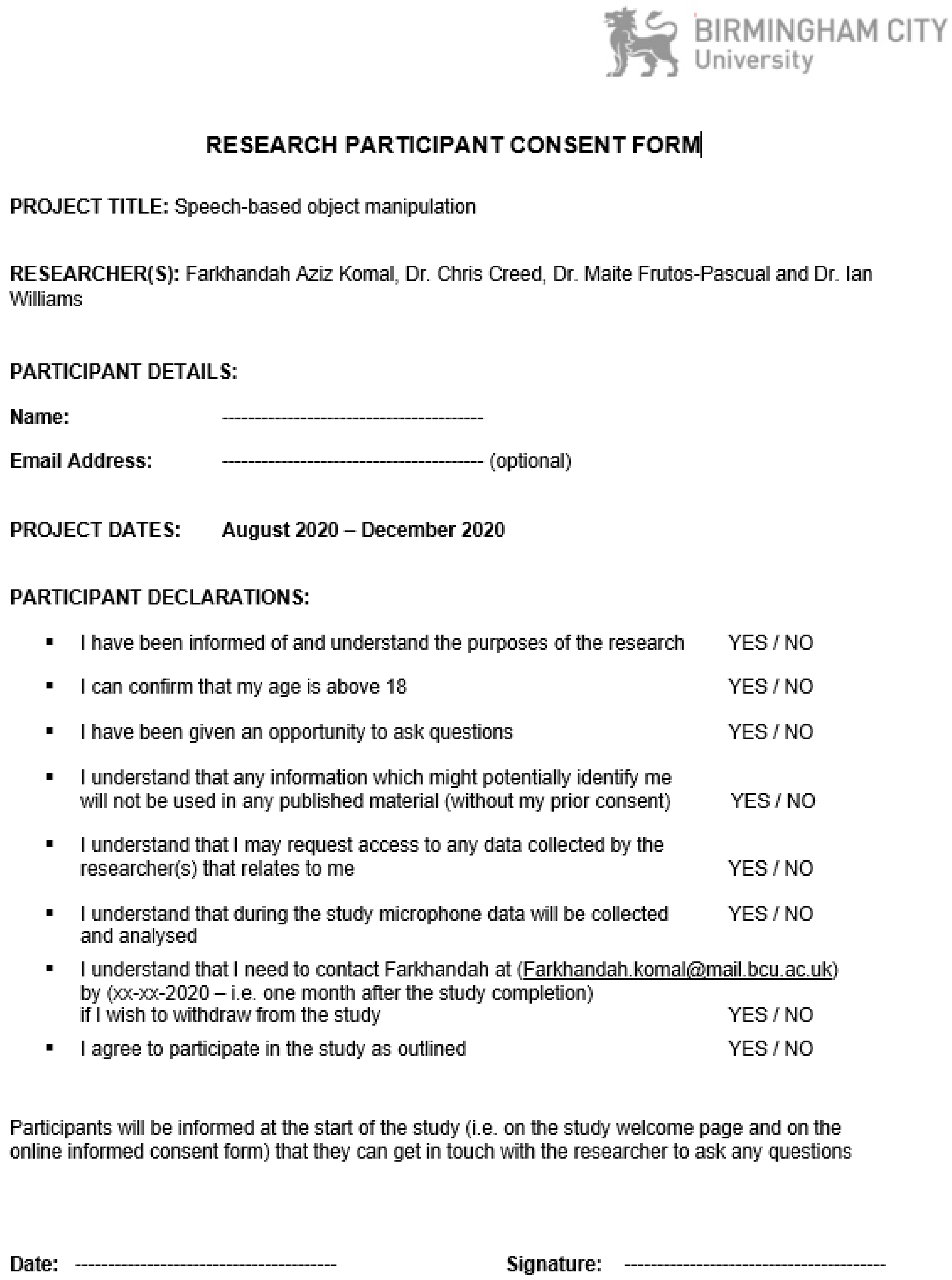
BENEFITS TO THE PARTICIPANTS
During this study you will have the opportunity to use different novel techniques for moving objects around a design canvas using voice commands. These interaction techniques can support people with

physical impairments in graphic design activity that is, moving objects around design canvas. All data collected will also be used to inform the design of future related research studies in this area.

RESEARCHER CONTACT DETAILS
Please contact **Farkhandah (Farkhandah.komal@mail.bcu.ac.uk)** if you have any further questions after completing this study.

COMPLAINTS
Please contact BCU_ethics@bcu.ac.uk if you wish to make a complaint.

Participants Consent Form:



BIRMINGHAM CITY
University

RESEARCH PARTICIPANT CONSENT FORM

PROJECT TITLE: Speech-based object manipulation

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos-Pascual and Dr. Ian Williams

PARTICIPANT DETAILS:

Name: --

Email Address: -- (optional)

PROJECT DATES: **August 2020 – December 2020**

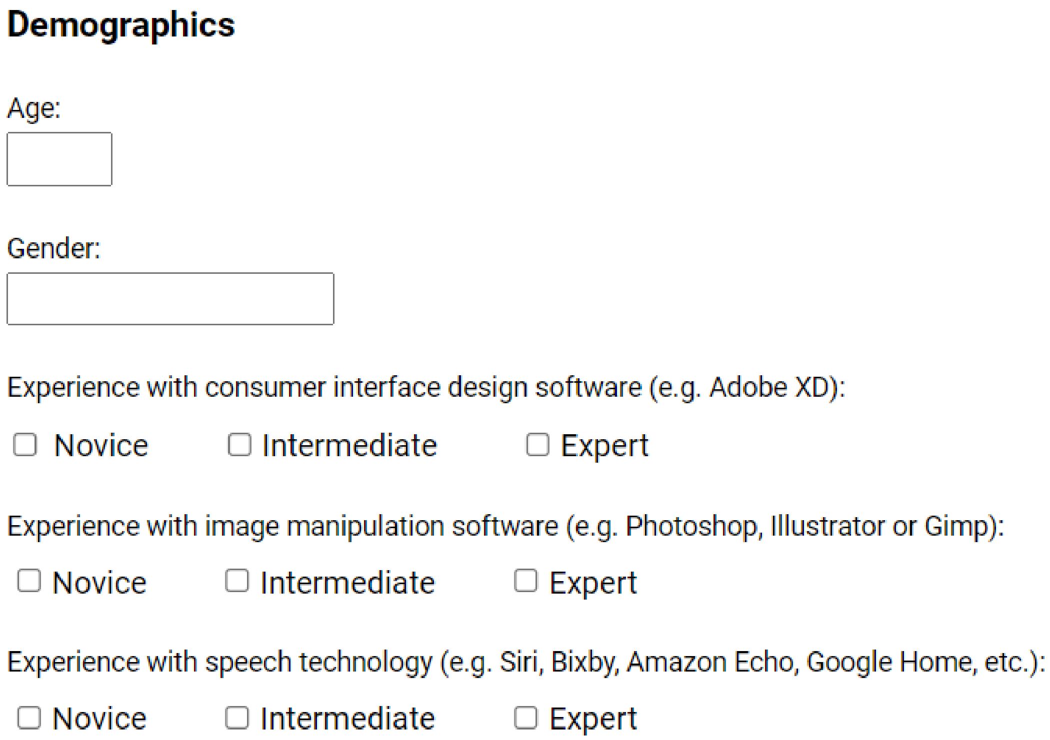
PARTICIPANT DECLARATIONS:

- I have been informed of and understand the purposes of the research YES / NO
- I can confirm that my age is above 18 YES / NO
- I have been given an opportunity to ask questions YES / NO
- I understand that any information which might potentially identify me will not be used in any published material (without my prior consent) YES / NO
- I understand that I may request access to any data collected by the researcher(s) that relates to me YES / NO
- I understand that during the study microphone data will be collected and analysed YES / NO
- I understand that I need to contact Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk) by (xx-xx-2020 – i.e. one month after the study completion) if I wish to withdraw from the study YES / NO
- I agree to participate in the study as outlined YES / NO

Participants will be informed at the start of the study (i.e. on the study welcome page and on the online informed consent form) that they can get in touch with the researcher to ask any questions

Date: --   **Signature:** --

Pre-test Questions:



Demographics

Age:

Gender:

Experience with consumer interface design software (e.g. Adobe XD):

☐ Novice ☐ Intermediate ☐ Expert

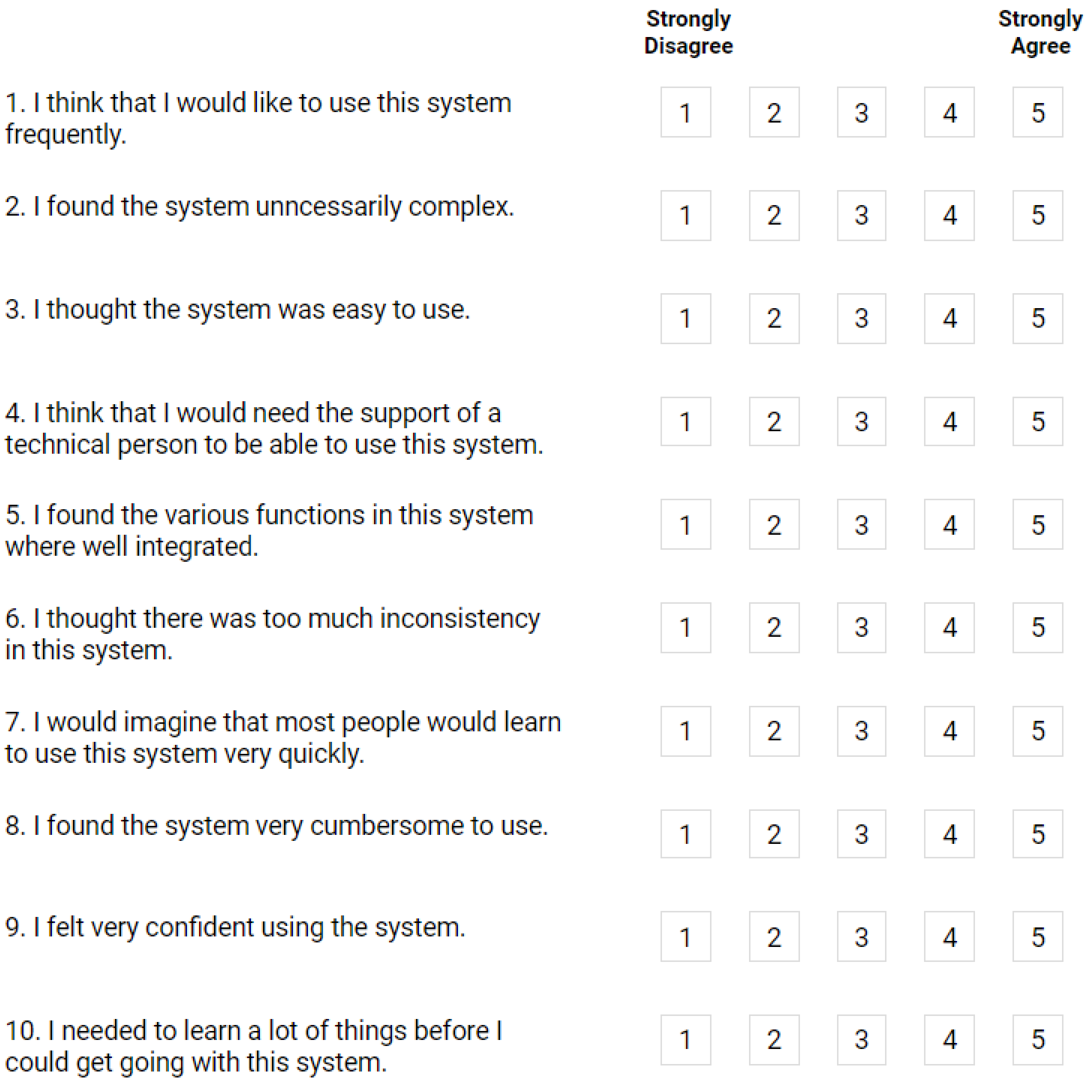
Experience with image manipulation software (e.g. Photoshop, Illustrator or Gimp):

☐ Novice ☐ Intermediate ☐ Expert

Experience with speech technology (e.g. Siri, Bixby, Amazon Echo, Google Home, etc.):

☐ Novice ☐ Intermediate ☐ Expert

System Usability Scale (SUS):



	Strongly Disagree				Strongly Agree
1. I think that I would like to use this system frequently.	1	2	3	4	5
2. I found the system unncessarily complex.	1	2	3	4	5
3. I thought the system was easy to use.	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5. I found the various functions in this system where well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this system.	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8. I found the system very cumbersome to use.	1	2	3	4	5
9. I felt very confident using the system.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Semi-structured Interview Questions (User Study 1):

1. What did you like about 'Speed Control'?
2. What did you dislike about 'Speech Control'?
3. What did you like about 'Location Guides'?
4. What did you dislike about 'Location Guides'?
5. What did you like about 'Positional Guides'?
6. What did you dislike about 'Positional Guides'?
7. Which interaction approach would you prefer the most for moving objects? (Speech Control, Location Guides, or Positional Guides)? And why?
8. What is your overall impression about the object positioning application?
9. How can we improve the overall functionality of object positioning application?

Open-Ended Questions (Follow-up Study):

1. What did you like about 'Location Guides'?
2. What did you dislike about 'Location Guides'?
3. What are your overall impressions about the Location Guides for object positioning?
4. How can we improve the overall functionality of object positioning application?

APPENDIX 8.2: OBJECT RESIZE (CHAPTER 4)

Ethical Approval:

BIRMINGHAM CITY
University

Faculty of Computing, Engineering & the Built Environment Research Office
Millennium Point, Curzon Street
Birmingham
B4 7XG

BCU_ethics@bcu.ac.uk

22/Apr/2021

Miss Farkhandah Aziz Komal

farkhandah.komal@mail.bcu.ac.uk

Dear Farkhandah Aziz ,

Re: Komal /#9315 /sub1 /R(B) /2021 /Apr /CEBE FAEC - Speech interaction for Object Resize Manipulation

Thank you for your application and documentation regarding the above activity. I am pleased to take Chair's Action and approve this activity with the following proviso:

- Please ensure that participants understand whether their voice commands are being recorded, and if so, how they will be stored and used, and whey they will be deleted.

Provided that you are granted Permission of Access by relevant parties (meeting requirements as laid out by them), you may begin your activity.

I can also confirm that any person participating in the project is covered under the University's insurance arrangements.

Please note that ethics approval only covers your activity as it has been detailed in your ethics application. If you wish to make any changes to the activity, then you must submit an Amendment application for approval of the proposed changes.

Examples of changes include (but are not limited to) adding a new study site, a new method of participant recruitment, adding a new method of data collection and/or change of Project Lead.

Please also note that the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee should be notified of any serious adverse effects arising as a result of this activity.

If for any reason the Committee feels that the activity is no longer ethically sound, it reserves the right to withdraw its approval. In the unlikely event of issues arising which would lead to this, you will be consulted.

Keep a copy of this letter along with the corresponding application for your records as evidence of approval.

If you have any queries, please contact BCU_ethics@bcu.ac.uk

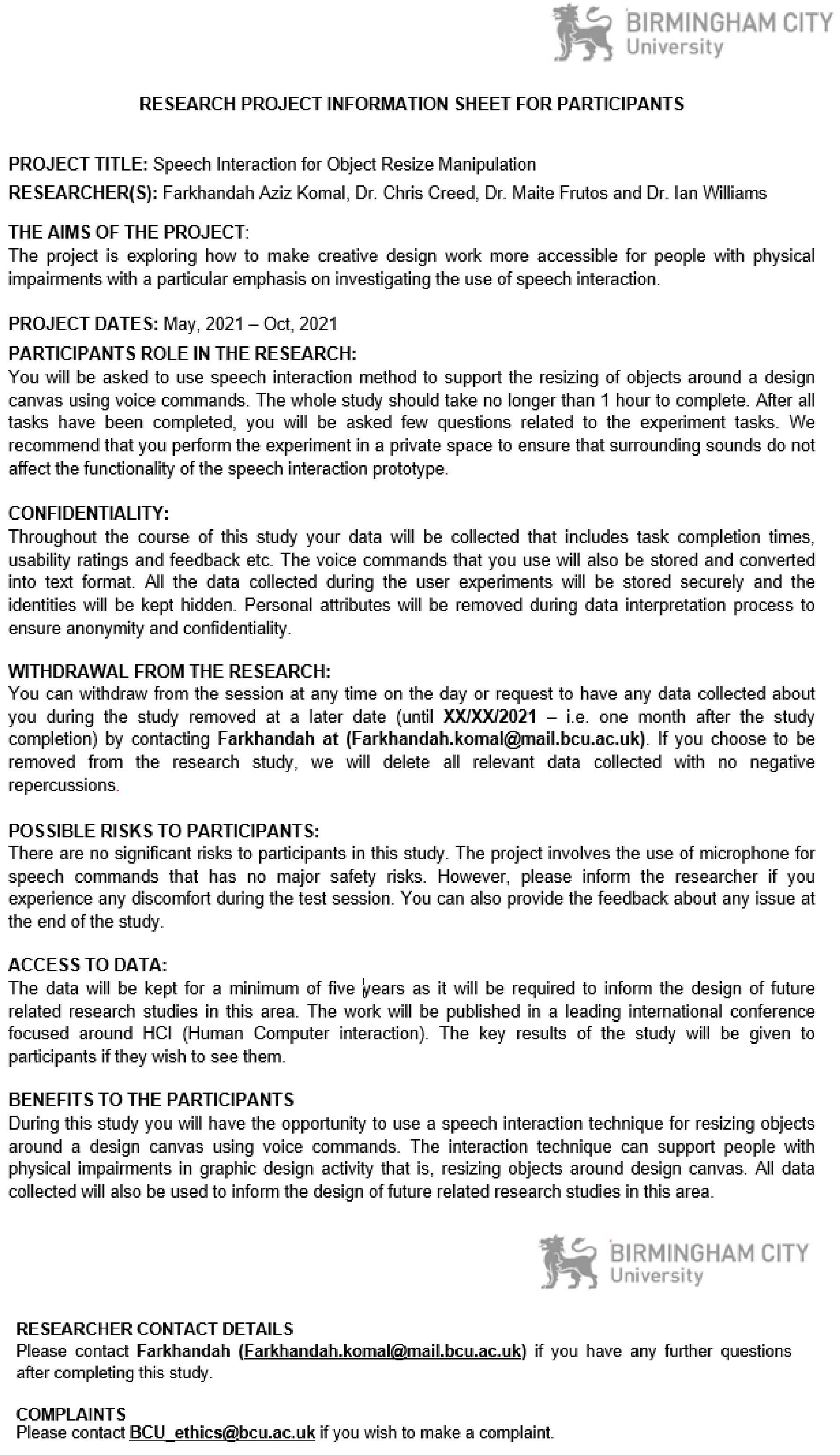
I wish you every success with your activity.

Yours Sincerely,

Professor Sharon Cox

On behalf of the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee

Participants Information Sheet:



BIRMINGHAM CITY University

RESEARCH PROJECT INFORMATION SHEET FOR PARTICIPANTS

PROJECT TITLE: Speech Interaction for Object Resize Manipulation

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos and Dr. Ian Williams

THE AIMS OF THE PROJECT:
The project is exploring how to make creative design work more accessible for people with physical impairments with a particular emphasis on investigating the use of speech interaction.

PROJECT DATES: May, 2021 – Oct, 2021

PARTICIPANTS ROLE IN THE RESEARCH:
You will be asked to use speech interaction method to support the resizing of objects around a design canvas using voice commands. The whole study should take no longer than 1 hour to complete. After all tasks have been completed, you will be asked few questions related to the experiment tasks. We recommend that you perform the experiment in a private space to ensure that surrounding sounds do not affect the functionality of the speech interaction prototype.

CONFIDENTIALITY:
Throughout the course of this study your data will be collected that includes task completion times, usability ratings and feedback etc. The voice commands that you use will also be stored and converted into text format. All the data collected during the user experiments will be stored securely and the identities will be kept hidden. Personal attributes will be removed during data interpretation process to ensure anonymity and confidentiality.

WITHDRAWAL FROM THE RESEARCH:
You can withdraw from the session at any time on the day or request to have any data collected about you during the study removed at a later date (until **XX/XX/2021** – i.e. one month after the study completion) by contacting **Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk)**. If you choose to be removed from the research study, we will delete all relevant data collected with no negative repercussions.

POSSIBLE RISKS TO PARTICIPANTS:
There are no significant risks to participants in this study. The project involves the use of microphone for speech commands that has no major safety risks. However, please inform the researcher if you experience any discomfort during the test session. You can also provide the feedback about any issue at the end of the study.

ACCESS TO DATA:
The data will be kept for a minimum of five years as it will be required to inform the design of future related research studies in this area. The work will be published in a leading international conference focused around HCI (Human Computer interaction). The key results of the study will be given to participants if they wish to see them.

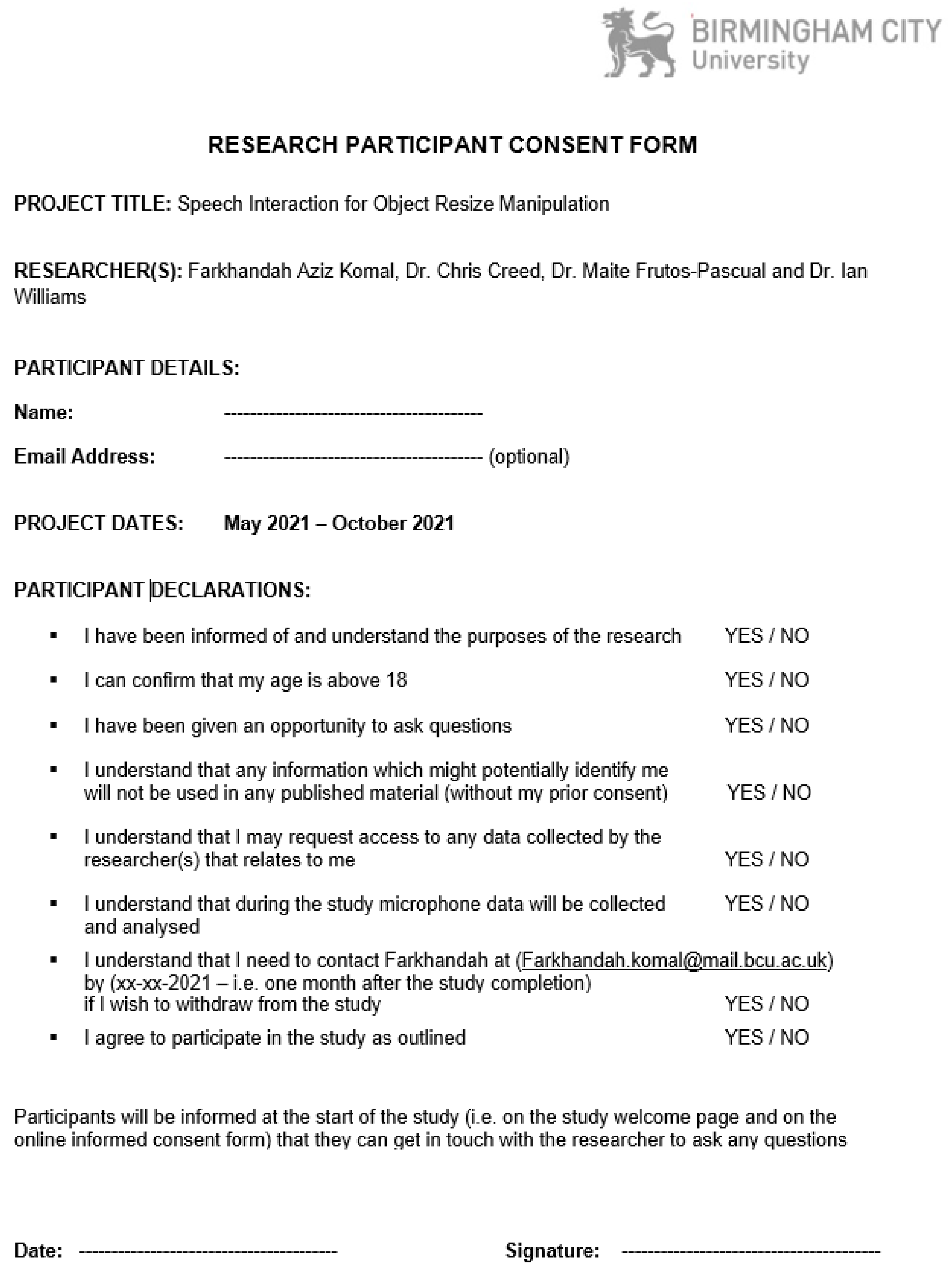
BENEFITS TO THE PARTICIPANTS
During this study you will have the opportunity to use a speech interaction technique for resizing objects around a design canvas using voice commands. The interaction technique can support people with physical impairments in graphic design activity that is, resizing objects around design canvas. All data collected will also be used to inform the design of future related research studies in this area.

BIRMINGHAM CITY University

RESEARCHER CONTACT DETAILS
Please contact **Farkhandah (Farkhandah.komal@mail.bcu.ac.uk)** if you have any further questions after completing this study.

COMPLAINTS
Please contact **BCU_ethics@bcu.ac.uk** if you wish to make a complaint.

Consent Form:



BIRMINGHAM CITY University

RESEARCH PARTICIPANT CONSENT FORM

PROJECT TITLE: Speech Interaction for Object Resize Manipulation

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos-Pascual and Dr. Ian Williams

PARTICIPANT DETAILS:

Name: --

Email Address: -- (optional)

PROJECT DATES: **May 2021 – October 2021**

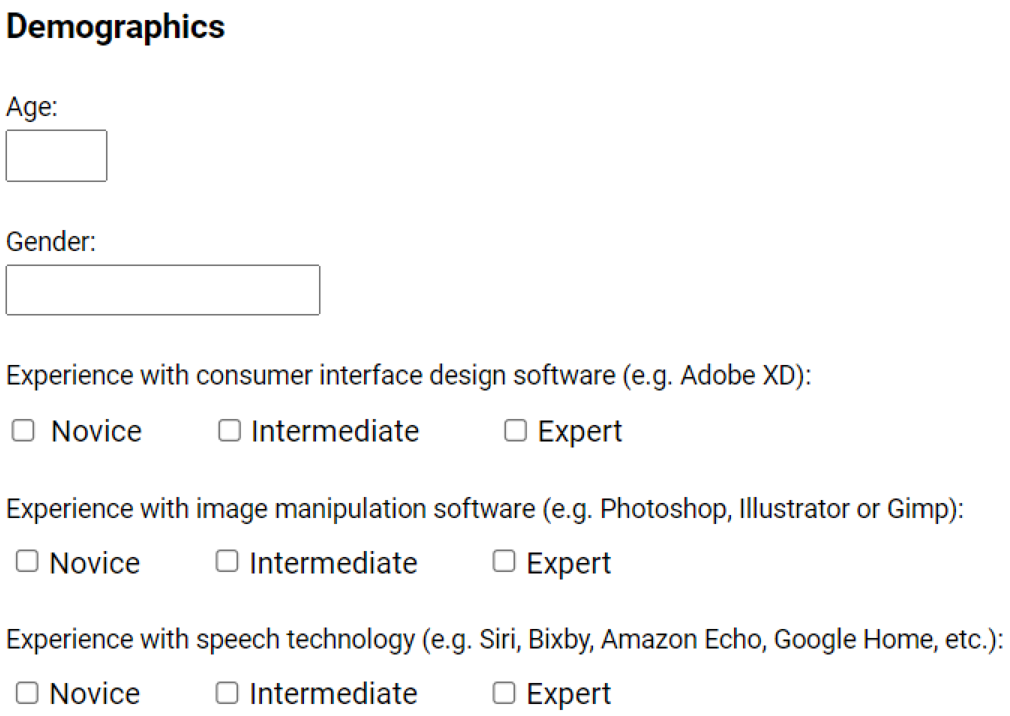
PARTICIPANT DECLARATIONS:

- I have been informed of and understand the purposes of the research YES / NO
- I can confirm that my age is above 18 YES / NO
- I have been given an opportunity to ask questions YES / NO
- I understand that any information which might potentially identify me will not be used in any published material (without my prior consent) YES / NO
- I understand that I may request access to any data collected by the researcher(s) that relates to me YES / NO
- I understand that during the study microphone data will be collected and analysed YES / NO
- I understand that I need to contact Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk) by (xx-xx-2021 – i.e. one month after the study completion) if I wish to withdraw from the study YES / NO
- I agree to participate in the study as outlined YES / NO

Participants will be informed at the start of the study (i.e. on the study welcome page and on the online informed consent form) that they can get in touch with the researcher to ask any questions

Date: -- **Signature:** --

Pre-test Questions:



Demographics

Age:

Gender:

Experience with consumer interface design software (e.g. Adobe XD):

☐ Novice ☐ Intermediate ☐ Expert

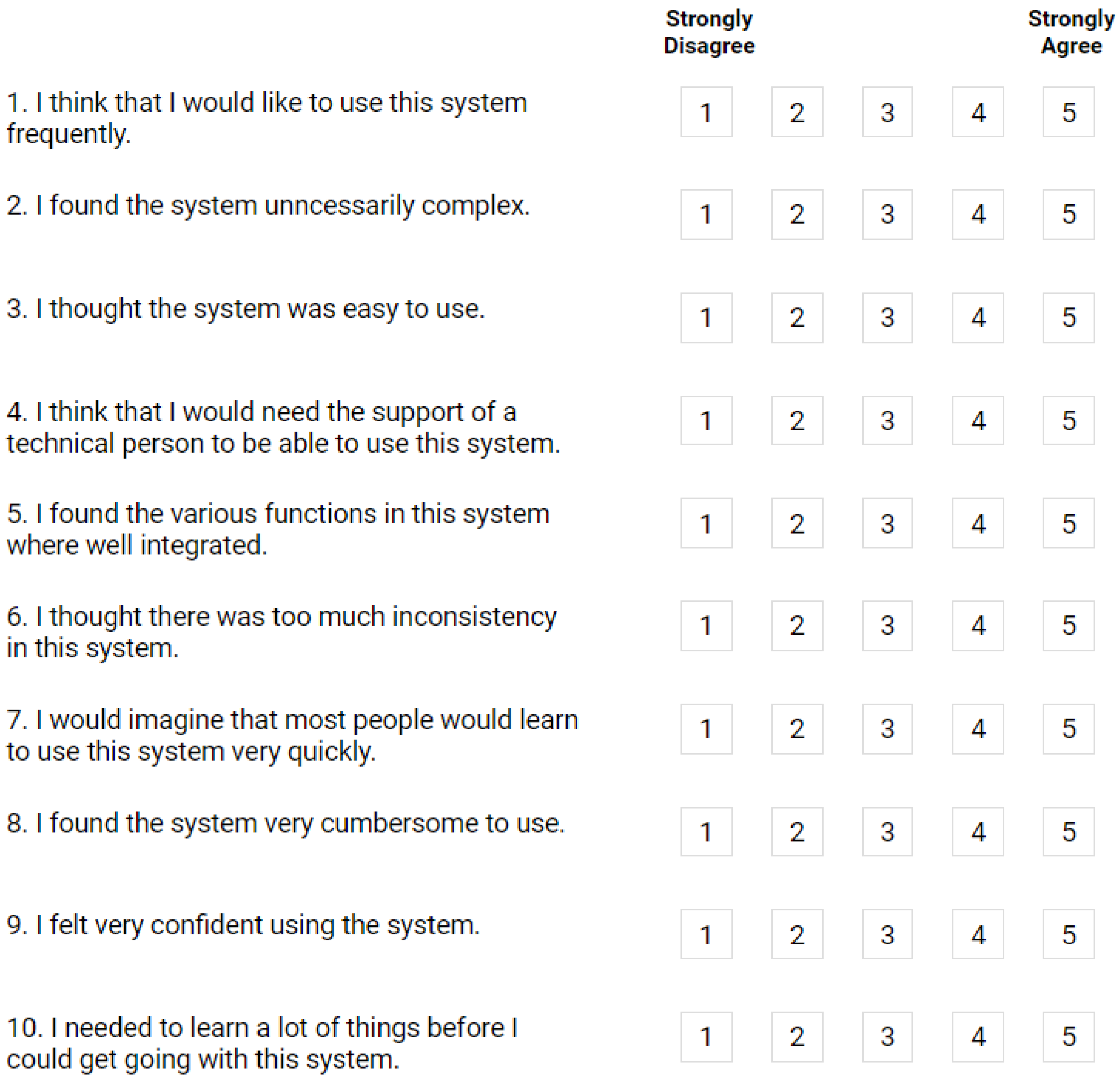
Experience with image manipulation software (e.g. Photoshop, Illustrator or Gimp):

☐ Novice ☐ Intermediate ☐ Expert

Experience with speech technology (e.g. Siri, Bixby, Amazon Echo, Google Home, etc.):

☐ Novice ☐ Intermediate ☐ Expert

Chapter 4: System Usability Scale (SUS):



	Strongly Disagree				Strongly Agree
1. I think that I would like to use this system frequently.	1	2	3	4	5
2. I found the system unncessarily complex.	1	2	3	4	5
3. I thought the system was easy to use.	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5. I found the various functions in this system where well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this system.	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8. I found the system very cumbersome to use.	1	2	3	4	5
9. I felt very confident using the system.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Chapter 4: Post study Open-ended Questions (Exploratory Study)

1. What did you like about using speech input for object resize tasks?
2. What did you dislike about using speech input for object resize tasks?
3. Are the presented speech commands institutive or do you have any alternative options?
4. What is your overall impression about the object resizing prototype?
5. How can we improve the overall functionality of object resizing prototype?

Chapter 4: Semi-structured Interview Questions (Main Resize Study)

1. What did you like about ‘NoSnap’?
2. What did you dislike about ‘NoSnap’?
3. What did you like about ‘UserSnap’?
4. What did you dislike about ‘UserSnap’?
5. What did you like about ‘AutoSnap’?
6. What did you dislike about ‘AutoSnap’?
7. Which interaction approach would you prefer the most for resizing objects? (NoSnap, UserSnap, AutoSnap)? And why?
8. What is your overall impression about the object resizing application?
9. How can we improve the overall functionality of object resizing application?

APPENDIX 8.3: OBJECT ROTATION (CHAPTER 5)

Ethical Approval:

Faculty of Computing, Engineering & the Built Environment Research Office
Millennium Point, Curzon Street
Birmingham
B4 7XG

BCU_ethics@bcu.ac.uk

21/Mar/2022

Miss Farkhandah Aziz Komal

farkhandah.komal@mail.bcu.ac.uk

Dear Farkhandah Aziz ,

Re: Komal /#10132 /sub2 /Am /2022 /Mar /CEBE FAEC - Speech Interaction for Object Rotation

Thank you for your application for approval of amendments regarding the above study. I am happy to take Chair's Action and approve these amendments.

Provided that you are granted Permission of Access by relevant parties (meeting requirements as laid out by them), you may continue your activity.

I can also confirm that any person participating in the project is covered under the University's insurance arrangements.

Please note that ethics approval only covers your activity as it has been detailed in your ethics application. If you wish to make any changes to the activity, then you must submit an Amendment application for approval of the proposed changes.

Examples of changes include (but are not limited to) adding a new study site, a new method of participant recruitment, adding a new method of data collection and/or change of Project Lead.

Please also note that the Committee should be notified of any serious adverse effects arising as a result of this activity.

If for any reason the Committee feels that the activity is no longer ethically sound, it reserves the right to withdraw its approval. In the unlikely event of issues arising which would lead to this, you will be consulted.

Keep a copy of this letter along with the corresponding application for your records as evidence of approval.

If you have any queries, please contact BCU_ethics@bcu.ac.uk

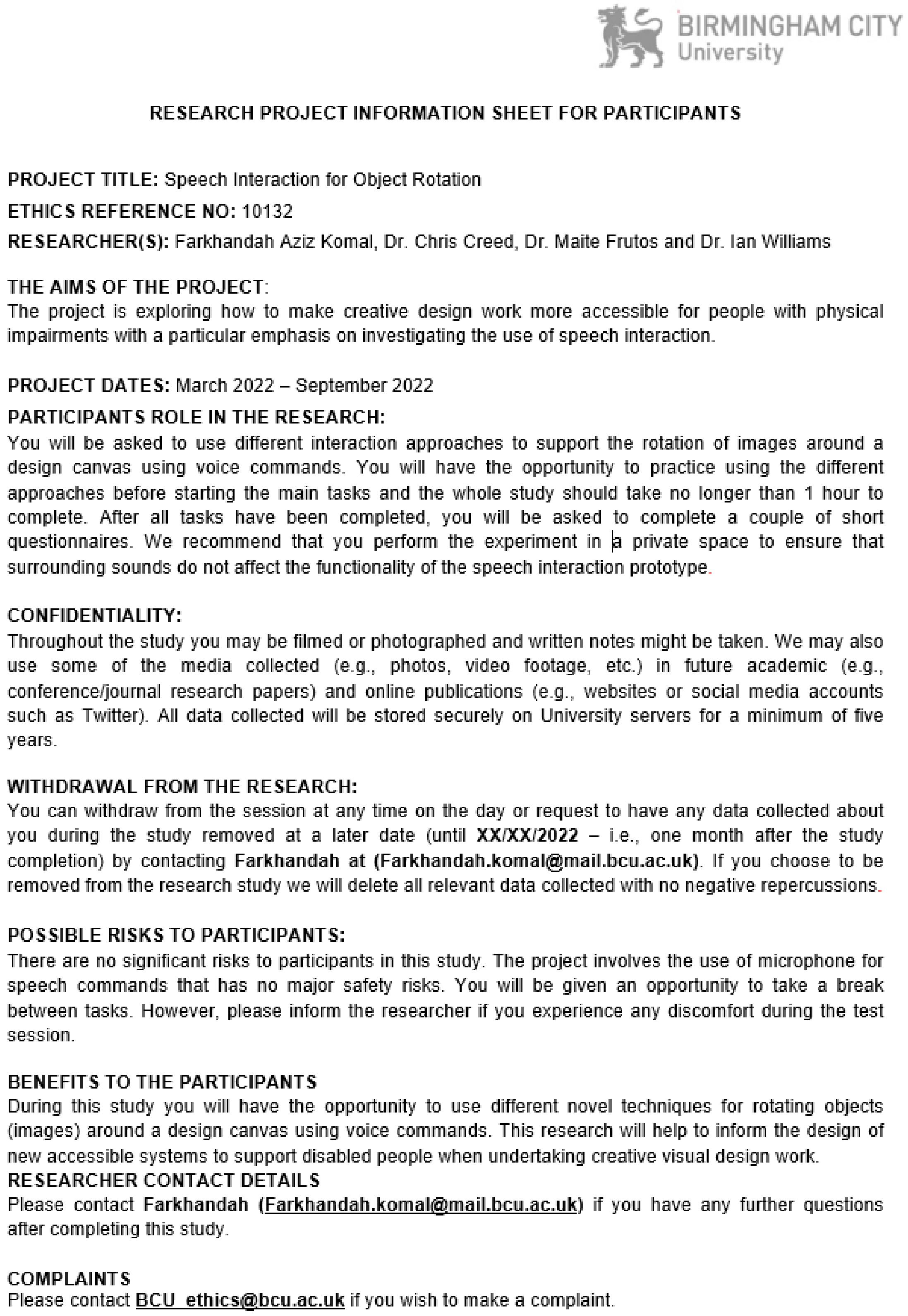
I wish you every success with your activity.

Yours Sincerely,

Professor Sharon Cox

On behalf of the Computing, Engineering and the Built Environment Faculty Academic Ethics Committee

Participants Information Sheet:



BIRMINGHAM CITY
University

RESEARCH PROJECT INFORMATION SHEET FOR PARTICIPANTS

PROJECT TITLE: Speech Interaction for Object Rotation

ETHICS REFERENCE NO: 10132

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos and Dr. Ian Williams

THE AIMS OF THE PROJECT:
The project is exploring how to make creative design work more accessible for people with physical impairments with a particular emphasis on investigating the use of speech interaction.

PROJECT DATES: March 2022 – September 2022

PARTICIPANTS ROLE IN THE RESEARCH:
You will be asked to use different interaction approaches to support the rotation of images around a design canvas using voice commands. You will have the opportunity to practice using the different approaches before starting the main tasks and the whole study should take no longer than 1 hour to complete. After all tasks have been completed, you will be asked to complete a couple of short questionnaires. We recommend that you perform the experiment in a private space to ensure that surrounding sounds do not affect the functionality of the speech interaction prototype.

CONFIDENTIALITY:
Throughout the study you may be filmed or photographed and written notes might be taken. We may also use some of the media collected (e.g., photos, video footage, etc.) in future academic (e.g., conference/journal research papers) and online publications (e.g., websites or social media accounts such as Twitter). All data collected will be stored securely on University servers for a minimum of five years.

WITHDRAWAL FROM THE RESEARCH:
You can withdraw from the session at any time on the day or request to have any data collected about you during the study removed at a later date (until **XX/XX/2022** – i.e., one month after the study completion) by contacting **Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk)**. If you choose to be removed from the research study we will delete all relevant data collected with no negative repercussions.

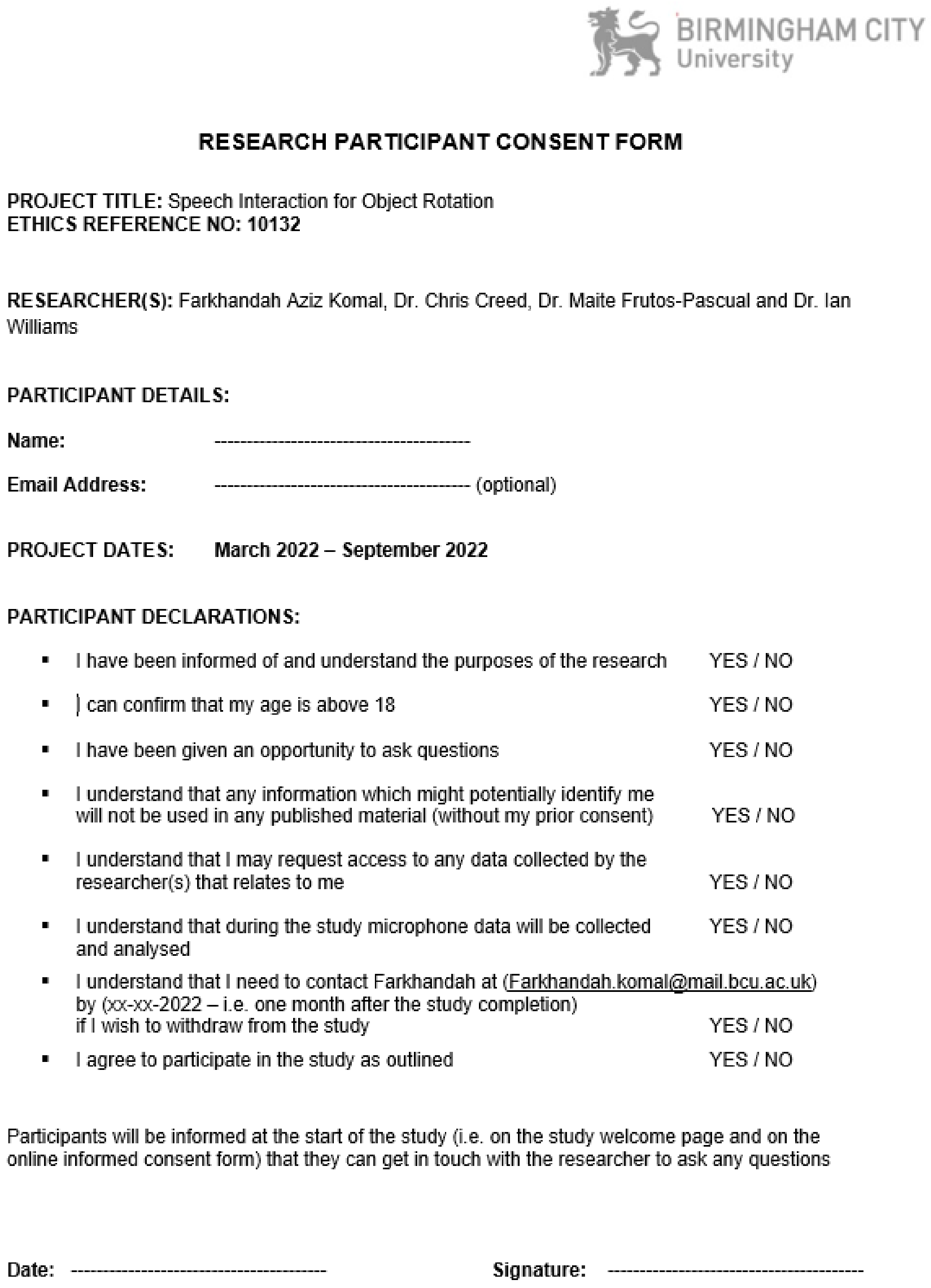
POSSIBLE RISKS TO PARTICIPANTS:
There are no significant risks to participants in this study. The project involves the use of microphone for speech commands that has no major safety risks. You will be given an opportunity to take a break between tasks. However, please inform the researcher if you experience any discomfort during the test session.

BENEFITS TO THE PARTICIPANTS
During this study you will have the opportunity to use different novel techniques for rotating objects (images) around a design canvas using voice commands. This research will help to inform the design of new accessible systems to support disabled people when undertaking creative visual design work.

RESEARCHER CONTACT DETAILS
Please contact **Farkhandah (Farkhandah.komal@mail.bcu.ac.uk)** if you have any further questions after completing this study.

COMPLAINTS
Please contact **BCU_ethics@bcu.ac.uk** if you wish to make a complaint.

Consent Form



RESEARCH PARTICIPANT CONSENT FORM

PROJECT TITLE: Speech Interaction for Object Rotation
ETHICS REFERENCE NO: 10132

RESEARCHER(S): Farkhandah Aziz Komal, Dr. Chris Creed, Dr. Maite Frutos-Pascual and Dr. Ian Williams

PARTICIPANT DETAILS:

Name: --

Email Address: -- (optional)

PROJECT DATES: **March 2022 – September 2022**

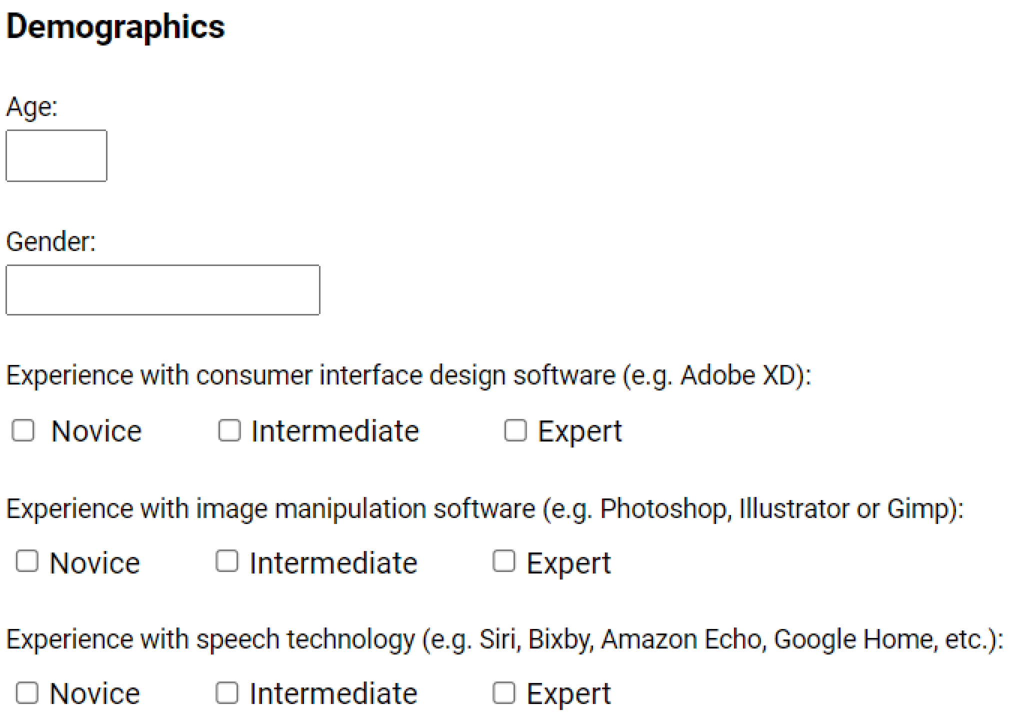
PARTICIPANT DECLARATIONS:

- I have been informed of and understand the purposes of the research YES / NO
- I can confirm that my age is above 18 YES / NO
- I have been given an opportunity to ask questions YES / NO
- I understand that any information which might potentially identify me will not be used in any published material (without my prior consent) YES / NO
- I understand that I may request access to any data collected by the researcher(s) that relates to me YES / NO
- I understand that during the study microphone data will be collected and analysed YES / NO
- I understand that I need to contact Farkhandah at (Farkhandah.komal@mail.bcu.ac.uk) by (xx-xx-2022 – i.e. one month after the study completion) if I wish to withdraw from the study YES / NO
- I agree to participate in the study as outlined YES / NO

Participants will be informed at the start of the study (i.e. on the study welcome page and on the online informed consent form) that they can get in touch with the researcher to ask any questions

Date: -- **Signature:** --

Pre-test Questions:



Demographics

Age:

Gender:

Experience with consumer interface design software (e.g. Adobe XD):

☐ Novice ☐ Intermediate ☐ Expert

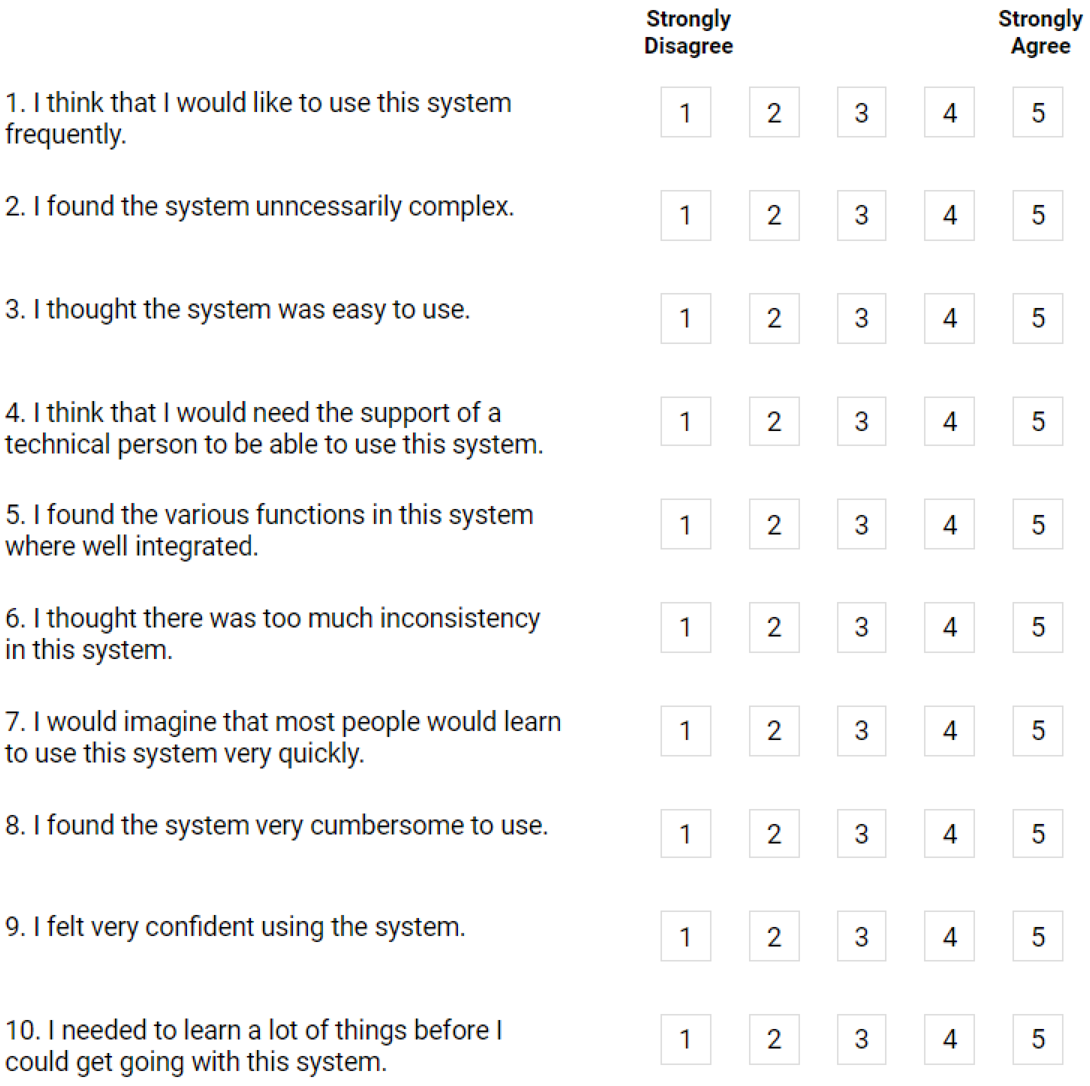
Experience with image manipulation software (e.g. Photoshop, Illustrator or Gimp):

☐ Novice ☐ Intermediate ☐ Expert

Experience with speech technology (e.g. Siri, Bixby, Amazon Echo, Google Home, etc.):

☐ Novice ☐ Intermediate ☐ Expert

Chapter 5: System Usability Scale (SUS):



	Strongly Disagree				Strongly Agree
1. I think that I would like to use this system frequently.	1	2	3	4	5
2. I found the system unncessarily complex.	1	2	3	4	5
3. I thought the system was easy to use.	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system.	1	2	3	4	5
5. I found the various functions in this system where well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this system.	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly.	1	2	3	4	5
8. I found the system very cumbersome to use.	1	2	3	4	5
9. I felt very confident using the system.	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system.	1	2	3	4	5

Chapter 5: Elicitation Study

1. Why did you choose that (certain) command for rotating objects?
2. Do you have any other alternative commands other than you used for rotating objects?
3. Do you think you would choose different commands if different types of objects are presented each time?
4. What are your overall impressions on using speech for object rotation tasks?

Chapter 5: Post-study Questions (Exploratory Study):

1. What did you like about using speech input for object rotation activity?
2. What did you dislike about using speech input for object rotation activity?
3. What is your overall impression about the object rotation prototype?
4. How can we improve the overall functionality of object rotation prototype?

Chapter 5: Semi-structured Interview Questions (Main Rotation Study):

1. What did you like about ‘Baseline Rotation’?
2. What did you dislike about ‘Baseline Rotation’?
3. What did you like about ‘Animation Rotation’?
4. What did you dislike about ‘Animation Rotation’?
5. What did you like about ‘Fixed-Jumps Rotation’?
6. What did you dislike about ‘Fixed-Jumps Rotation’?
7. Which interaction approach would you prefer the most for rotating objects? (Baseline Rotation, Animation Rotation, Fixed-Jumps Rotation)? And why?
8. What is your overall impression about the object rotation application?
9. How can we improve the overall functionality of object rotation application?