Universal Access in the Information Society
https://doi.org/10.1007/s10209-025-01212-8

LONG PAPER

Inclusive speech interaction techniques for creative object rotation

Farkhandah Aziz[1] · Chris Creed[1] · Maite Frutos-Pascual[1] · Ian Williams[1]

Accepted: 5 March 2025

Abstract
Speech interaction holds significant potential to make creative visual design activities more inclusive for people with physical impairments, although no work has yet investigated the feasibility of graphical object rotation via voice control. An elicitation study with disabled participants ($N=12$) is initially presented where candidate voice commands for rotation actions are identified. The use of these commands is then evaluated in an exploratory study with people who have physical impairments ($N=12$). Results found all participants could successfully complete a series of rotation tasks, although interaction issues were also identified (e.g., estimating rotation transformation angles). To further investigate these challenges, three different voice-controlled rotation approaches were developed: Baseline-Rotation, Fixed-Jumps, and Animation-Rotation. These methods were evaluated with disabled participants ($N=25$) with results highlighting that all three approaches supported users in successfully rotating graphical objects, although Animation-Rotation was found to be more efficient and usable than the other methods.

Keywords Assistive technology · Accessibility · Speech interaction · Object manipulation · Inclusive visual design · Creative design

1 Introduction

Traditional input devices such as a mouse and keyboard present significant challenges for people with physical impairments (i.e., affecting upper body limbs such as hands, arms, or shoulders) in producing creative work using mainstream visual design applications (e.g., Figma, Photoshop, Illustrator, and XD [1–4]). Speech input is an alternative interaction modality that holds potential in making creative design work more accessible [29, 30, 33], although there has been limited research around this area to date. Previous research has started to investigate the use of speech interaction to support creative design work [7, 17–19, 27, 36] with initial work demonstrating the possibilities associated with this approach. In particular, studies have recently examined voice-controlled object transformation methods such as positioning and movement of digital assets around a design canvas [9, 15, 21], as well as the resizing of graphical objects [8, 26, 36]. The rotation of digital objects is another essential transformation approach that is widely and regularly used by designers. This is typically facilitated in mainstream creative applications through manipulating small transformation handles (attached to an object) via mouse dragging movements.æ However, there has been no empirical research to date exploring the viability of 2D digital object rotation via speech interaction to support people with physical impairments. It therefore remains unclear how this common form of object transformation can be supported via voice control and what the optimal approaches are for enabling efficient and accurate rotation of digital assets.

We address the lack of current work in this area through presenting three separate research studies—the first is an elicitation study with disabled participants identifying candidate voice commands that can be used to facilitate rotation transformations. These commands were integrated into an interactive prototype that was evaluated in an exploratory study with people who have physical impairments. Results found that all participants could successfully complete a

✉ Farkhandah Aziz
Farkhandah.komal@mail.bcu.ac.uk

Chris Creed
chris.creed@bcu.ac.uk

Maite Frutos-Pascual
maite.frutos@bcu.ac.uk

Ian Williams
ian.williams@bcu.ac.uk

[1] DMT Lab, Birmingham City University, Birmingham, UK

Published online: 25 March 2025

series of common rotation tasks, although key interaction issues were also identified (e.g., estimating rotation transformation dimensions). To further investigate these challenges, three different voice-controlled rotation approaches were developed: Baseline-Rotation, Fixed-Jumps, and Animation-Rotation. These methods were evaluated by disabled participants ($N=25$) with results highlighting that all three approaches supported users in successfully rotating graphical objects, although Animation-Rotation was found to be more efficient and usable than both Baseline-Rotation and Fixed-Jumps.

This work therefore presents five core contributions: (1) an elicitation study identifying candidate speech commands to support object rotation transformations on graphical assets, (2) an exploratory study investigating the feasibility of speech interaction for object rotation, (3) development of three speech interaction approaches facilitating digital object rotation, (4) user evaluations with participants who have physical impairments where new insights on the use of speech interaction for object rotation are identified, and (5) empirical evidence highlighting the efficacy and benefits of the Animation-Rotation approach for object rotation to support creative design activities.

2 Related work

2.1 Creative work using speech interaction

Previous literature exploring the potential of speech interaction to support creative design work has typically utilized multimodal approaches. For instance, in early work Hauptmann [20] evaluated a multimodal approach (combining speech and gesture) against different interaction approaches for moving, scaling, and rotating a cube via a limited vocabulary set (e.g., "left" and "right" commands). Results highlighted that speech could be used as part of a multimodal approach to manipulate digital cubes, although only a small set of manipulation operations were performed using a single cube shaped object. Pausch and Leatherby [33] also presented a graphical editor for basic drawing operations using a mouse and speech input. 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. The authors compared this multimodal approach against a standard mouse only interaction and found that the multimodal solution presented some interaction benefits (e.g., in terms of faster task completion times). Similarly, Gourdol et al. [16] also presented a multimodal interaction technique combining speech with mouse and keyboard inputs within their VoicePaint application. Drawing operations such as selecting paint brushes and sizes, adding shapes, and colour selection were handled by voice input while text input was performed using a keyboard. Hiyoshi and Shimazu [21] also presented a multimodal approach to support object positioning where mouse pointing was used to specify a target position for a basic shapes and speech input was used to complete movements (e.g., via statements such as "place the object here"). Furthermore, Nishimoto et al. [30] investigated a similar multimodal speech technique and compared against non-speech approaches (i.e., a mouse and keyboard) where results found the multimodal method to be easier to use and learn for producing creative work.

Sedivy and Johnson [36] investigated a multimodal approach using speech input with a tablet and stylus to support sketching activities. The system utilized various speech commands for operations such as colouring, grouping, layering, scaling, and resizing, alongside the use of a stylus for drawing. A user evaluation found that speech commands supported the creative process of participants, reduced tool selection time and cognitive processing. Alsuraihi and Rigas [7] compared a multimodal speech approach (i.e., a combination of voice recognition, mouse and keyboard) with a traditional mouse and keyboard. The system was evaluated across a range of standard design activities (e.g., creating buttons, choosing colours, writing text, and the selection of different tools) with results highlighting that voice recognition reduced reliance on traditional inputs for selection of drawing tools (thus presenting interaction benefits over the use of only a mouse and keyboard).

Moreover, Van der Kamp and Sundstedt [41] examined the use of voice input with eye gaze where voice commands ("start, "stop", "snap", "open colours", etc.) were used for drawing shapes and eye gaze for positioning the mouse cursor. Results highlighted that a gaze and speech combination supported a more efficient drawing process, as well presenting a more engaging experience for participants. Laput et al. [27] 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). Although 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 presented interaction benefits over a "touch-only" version of the application. Srinivasan et al. [40] also presented a multimodal approach using speech commands and touch input for image editing operations. 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 positive perceptions from participants although there were issues with speech recognition during the study.

Furthermore, Kim et al. [26] utilized 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 the creative practice of expert designers (e.g., "brush" to select the brush tool). A user evaluation found that these short voice commands helped creative experts with accessing 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 free-form drawings—for instance, Harada et al. [17–19] 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), although the authors highlighted interaction challenges with this approach in relation to smooth mapping of vowel sounds to the movements of a brush tool.

2.2 Object manipulation using speech interaction

Whilst the studies highlighted have all explored the potential to utilize voice control within creative domains, a fundamental area where there has been limited research to date is around how digital assets can be efficiently manipulated and transformed using this method of interaction. The previous studies highlighted have tended to examine basic approaches for positioning of objects through simple commands such as "left", "right", "up" and "down" [12, 25, 49]. Aziz et al. [9] developed this work further through exploring how the use of positional and location guides can support both rapid and accurate movement of objects across a design canvas. Similarly, in terms of resizing of objects, relatively simple approaches have been explored to date such as the use of "size 50", "butterfly brush size 10", "shrink", "enlarge", "bigger" and "smaller" commands to transform an object's properties [26, 43]. Aziz et al. [8] built on this initial work to examine the use of voice-controlled alignment guides and snapping features to support the resizing of objects.

However, whilst initial work has started to explore the viability of positioning and resizing objects, there has been much less focus around object rotation. Previous work has investigated object rotation using multimodal input approaches that utilize speech interaction—for example, Williams and Ortega 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 [43, 44]. Similarly, Alibay et al. [6] utilized the combination of speech and gestures for the selection and rotation of objects within a 3D digital design context. Furthermore, House et al. [22] 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). These studies demonstrate the potential of voice control to support object rotation, although there has been no empirical work investigating the rotation of objects using speech commands as the primary input modality within a creative 2D visual design context. It therefore remains unclear whether voice-controlled approaches investigated in previous studies for other object transformations (e.g., positioning of assets via commands such as "left" and "right") are also appropriate in this context or whether they present new interaction issues (e.g., users potentially experiencing challenges in determining transformation angles when working with degrees as opposed to pixels). We address the limited research in this area through three research studies investigating how object rotation can be facilitated via speech interaction for people with physical impairments affecting upper body limbs. This is a fundamental transformation activity that designers need to regularly perform, although it remains unclear which voice commands are most appropriate to support this activity and which interaction challenges may need to be overcome to facilitate the effective rotation of objects.

3 Study 1 (elicitation study)—voice commands for object rotation

Given the limited work around rotating objects via speech interaction within a 2D design space, it is 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 [45]. 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 (Fig. 1-i.e., images, text, lines, and icons) where users had to verbalize the commands they would use, to rotate the objects to relevant target placeholders. Each type of object (i.e., image, text, line, and icon) involved four tasks where two objects were rotated in a clockwise direction and another two were rotated in anti-clockwise directions. Furthermore, two tasks among each type of object were rotated with "smaller" transformations and two were rotated with "larger" transformations (based on the angle of target placeholders). The larger and smaller threshold transformations were informed by earlier research investigating rotation techniques [28, 34] where large transformations were classified as being up to 180 degrees (between 70 to 180) and small transformations were considered to be between 20 to 60 degrees. Since there are no formal standards around the classification of rotation transformation angles, these studies were used as a guide in

Fig. 1 Four different types of objects (image, text, line, icon) used for object rotation during elicitation and exploratory

our work to define larger transformations as having a rotation angle of 180 degrees or above and smaller transformations as less than or equal to 50 degrees. The variety of objects and range of transformations were presented as referents to ensure participants performed a variety of activities whilst verbalizing vocal commands.

3.1 Methodology

3.1.1 Participants

Twelve participants with physical impairments (8 male and 4 female) were recruited through online advertisements and via existing links. 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, their experience with interface prototyping, graphical manipulation applications, speech technology, and assistive tools. Eight participants identified as experiencing repetitive strain injury (RSI), two with motor neurone disease (MND), one with multiple sclerosis (MS), and one with tenosynovitis (Table 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 1.

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 [3] was used to create the prototype for the elicitation study (Fig. 2).

3.1.3 Procedure

Institutional ethical review board approval was initially obtained for this study. During the testing session, the researcher shared their screen content where they displayed the overview of the study to participants, followed by obtaining consent from participants, and asking pre-test questions (i.e., 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 were a total of sixteen actions which consisted of four different object types (i.e., images, text, lines, and icons—Fig. 1). Each type of object (i.e., image, text, line, and icon) 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 corresponding vocal command suggestions. The four different object types were evaluated (i.e., images, text, lines, and icons) to observe whether these common design assets influenced the types of commands participants issued. Whilst other objects could also have been tested (e.g., custom shapes), it was felt that the selection of these four common object types provided a balance between investigating voice commands across different asset types and ensuring users were not overloaded with too many tasks during the testing sessions.

Participants were provided with a single training task before being presented with the referents for each object type (i.e., images, text, lines, and icons). The training task consisted of a single object displayed on the blank canvas where participants were asked to verbalize the speech

Table 1 Participants details: age, gender, physical impairment and condition details

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;	IP: Average; GM: 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

GM, Graphical Manipulation (software); IP, Interface Prototyping (Software); ST, Speech Technology; AT, Assistive Technology

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 (Fig. 2). When completing tasks, participants were not instructed on whether the rotation had to be completed in a particular direction (e.g., left/right or clockwise/anti-clockwise) to avoid biasing their responses. Standard properties of the interaction object (i.e., width, height, xy positions, and current rotation angle) were also displayed in the right sidebar of the interface.

Referents were displayed in a randomized order to minimize the potential for order bias. Participants started by verbalizing 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 (e.g., either icons, text, or lines). 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 min and were video recorded for later analysis.

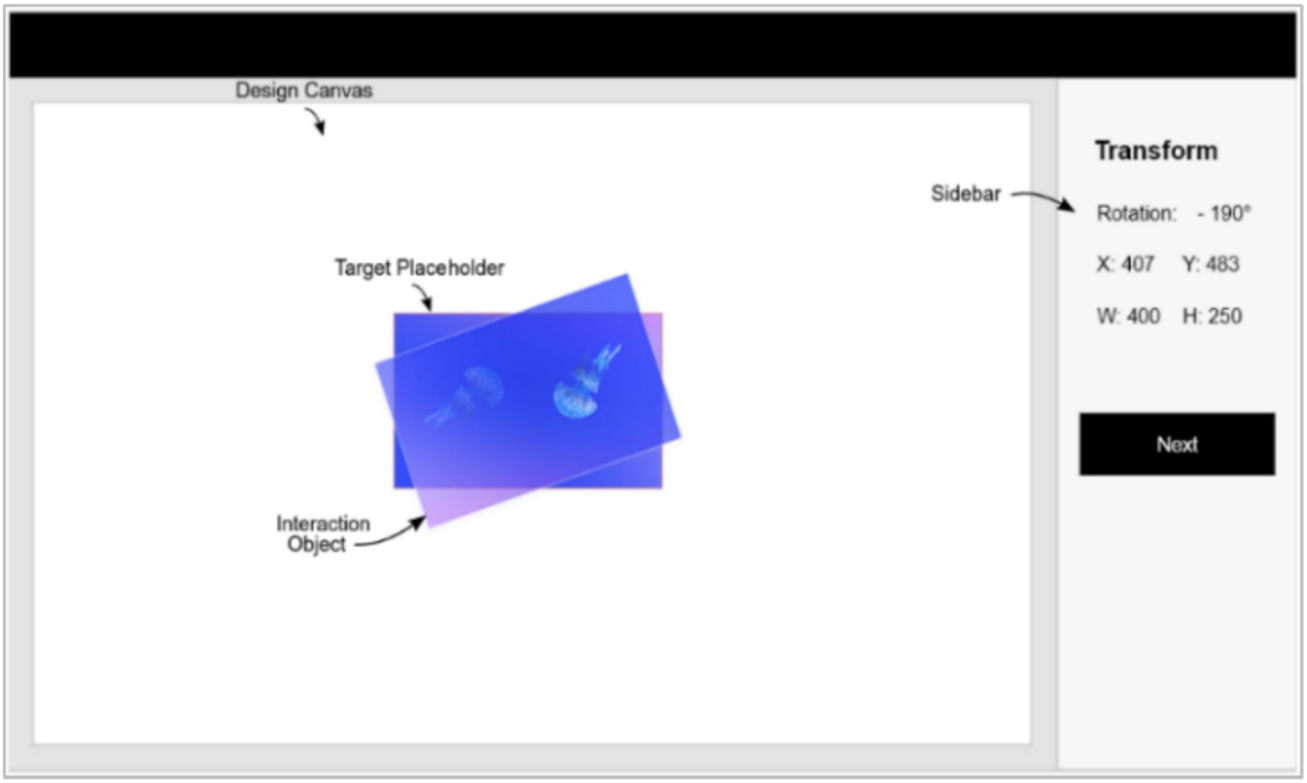

Fig. 2 Elicitation study prototype interface showing creative assets (a design canvas, an interaction object, a target placeholder, and a sidebar consisting of attributes of the interaction object)

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": *"…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"* (P12). 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 utilized clockwise/anticlockwise commands during first set of tasks (e.g., images), they then utilized 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 2).

In relation to objects with large transformations, nine participants suggested that they would repeat the same voice commands to get an object to their desired rotation position. For instance, if an object did not reach a target position after issuing a "left 180" command, they would then 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 distances—*"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'"* (P6). Participants did not highlight any issues related to objects requiring smaller transformations as they felt it would be easy to rotate these objects with a single or relatively few follow-up commands. A key theme emphasized 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). P8 also suggested that they would consider a customized set of commands (as supported by Dragon software) comprised

Table 2 Elicitation study—voice commands highlighted by each participant

ID	Frequently used commands	Other suggested commands
P1	Clockwise [xDegree], anti-clockwise [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

of "random" terms that are short and easy to pronounce (e.g., "alpha" or "bravo" for "right").

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 types (i.e., larger and smaller) 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.

4 Study 2 (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 [42] for speech recognition (Fig. 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 current rotation angle).

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"). The supported voice commands (Fig. 3g) are also visible at the bottom of the design canvas to help users in recalling the available commands. Switch input (e.g., a keyboard, mechanical switch, head tracker, or a foot pedal, etc.) could also be utilized to initiate the speech recognizer 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 recognized. Figure 3 shows an example where an interaction object has been rotated into a clockwise direction using voice command "right 25".

4.1 Methodology

4.1.1 Participants

Twelve participants with physical impairments (6 female and 6 male) were recruited through online advertisements using social media and existing network links. Participants were aged between 22 to 50 years (M = 33.75, SD = 7.47) and all were native English speakers. Table 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.

4.1.2 Apparatus

The testing sessions were conducted online using Zoom or Microsoft Teams depending on each participant's preference.

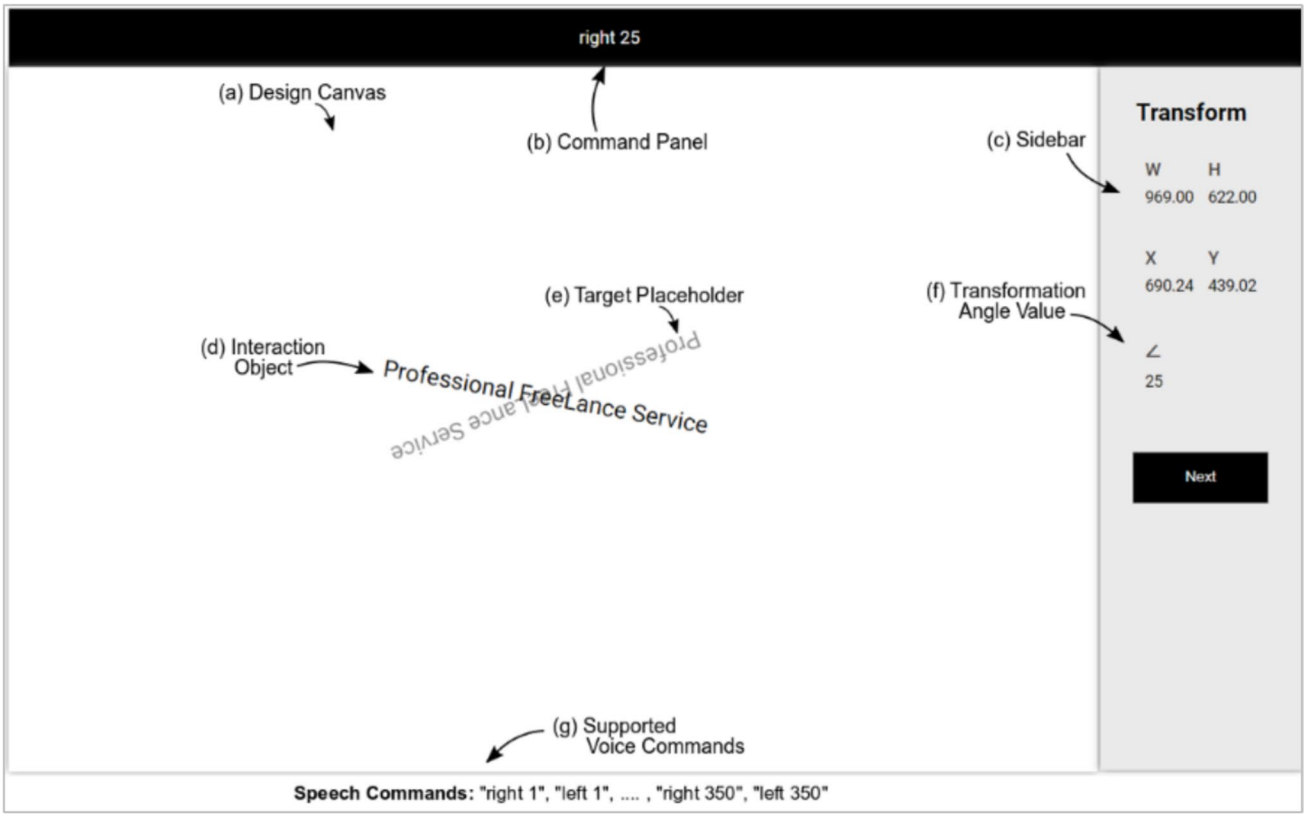

Fig. 3 Exploratory study research prototype consisting of **a** design canvas, **b** command panel, **c** sidebar, **d** interaction object, **e** target placeholder, **f** transformation angle value, and **g** supported voice commands

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 recognizer. The study was conducted using the Google Chrome browser to ensure compatibility with the Web Speech API.

4.1.3 Procedure

Institutional ethical review board approval was obtained before conducting user experiments. During the testing session, participants were provided with an overview of the study followed by pre-test questions requesting demographic information, details around the nature of their physical impairments, as well as technical experience with interface prototyping, graphical manipulation software and speech technology. The same four types of objects used in the elicitation study were utilized 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) (Fig. 3). 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 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 transformations.

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 four tasks. The order of these tasks was counterbalanced to reduce the potential for order bias. At the end of the testing session, participants were administered the System Usability Scale (SUS) questionnaire and a post-study interview was conducted. Testing sessions were video recorded for

Table 3 Participant Details: Age, Physical Impairments, Condition Details, and Technical Experience

ID	Age/gender	Physical impairments	Condition details	Technical experience
1	42 (M)	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	IP Average; GM: 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

further analysis with all testing sessions lasting between 35 to 40 min.

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 a user issued their first voice command for a new task until they ceased interaction after uttering their final command. 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 categorized into three categories: "Speech Misrecognition" (where system incorrectly identified the commands), "System Errors" (due to latency issues with the Web Speech API), and "Unsupported commands" (where users issued commands unrelated to supported set of commands). SUS was used to evaluate overall perceptions of usability for the object rotation approach.

4.2 Results

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

4.2.1 Task completion time

The overall mean task completion time across all twelve participants was 10 min and 56 s (SD = 1.44), while the mean task completion time across all large transformation tasks was 1.65 min (SD = 0.59) and 0.98 min (SD = 0.49) for small transformations. In terms of image tasks, the mean task completion time across all tasks was 1.50 min (SD = 0.58), while for large transformations it was 1.83 min (SD = 0.57) and 1.17 min (SD = 0.38) for smaller rotation tasks. For text tasks, the mean task completion was 1.27 min (SD = 0.63) across all tasks, while for large transformations it was 1.60 min (SD = 0.57) and 0.94 min (SD = 0.50) for smaller rotation tasks. The mean task completion time across all line tasks was 1.03 min (SD = 0.54), while for large transformations it was 1.33 min (SD = 0.44) and 0.73 min (SD = 0.46)

Table 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)	Task completion time (Mean (M) = Seconds)	No. of speech commands (*n*), Mean (M), SD
Image	1	Large	0	180	M = 54.87 SD = 15.88	*n* = 96, M = 8.00 SD = 2.41
	2	Large	175	− 45	M = 55.32 SD = 19.51	*n* = 100, M = 8.33 SD = 2.65
	3	Small	− 10	5	M = 36.01 SD = 11.56	*n* = 57, M = 4.75 SD = 1.48
	4	Small	− 120	− 145	M = 34.68 SD = 11.44	*n* = 56, M = 4.66 SD = 1.31
Text	1	Large	− 60	130	M = 47.33 SD = 16.90	*n* = 91, M = 7.58 SD = 1.25
	2	Large	175	− 45	M = 48.96 SD = 17.34	*n* = 92, M = 7.68 SD = 1.70
	3	Small	15	35	M = 27.85 SD = 15.16	*n* = 53, M = 4.41 SD = 2.21
	4	Small	20	− 5	M = 28.70 SD = 15.20	*n* = 57, M = 4.75 SD = 2.01
Lines	1	Large	10	210	M = 40.75 SD = 13.15	*n* = 74, M = 6.16 SD = 1.28
	2	Large	115	− 75	M = 41.73 SD = 13.41	*n* = 80, M = 6.67 SD = 1.17
	3	Small	25	50	M = 21.33 SD = 13.48	*n* = 49, M = 4.08 SD = 0.86
	4	Small	− 90	− 125	M = 22.86 SD = 14.64	*n* = 54, M = 4.50 SD = 1.12
Icons	1	Large	60	275	M = 55.71 SD = 18.73	*n* = 90, M = 7.50 SD = 1.84
	2	Large	180	− 60	M = 53.68 SD = 18.74	*n* = 83, M = 6.91 SD = 1.89
	3	Small	− 45	− 35	M = 33.66 SD = 14.09	*n* = 69, M = 5.75 SD = 1.64
	4	Small	75	35	M = 32.28 SD = 14.93	*n* = 63, M = 5.25 SD = 1.78

for smaller tasks. In relation to icon tasks, the mean task completion time across all tasks was 1.46 min (SD = 0.66), while for large transformations it was 1.82 min (SD = 0.62) and 1.09 min (SD = 0.48) for smaller tasks. Mean task completion times across all transformations are presented in Table 4.

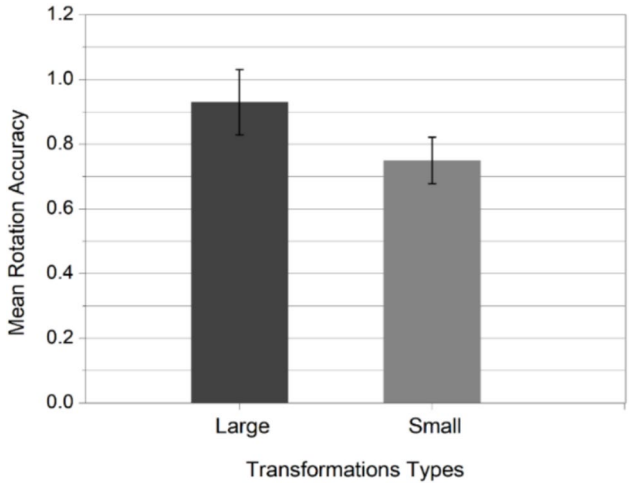

Fig. 4 Mean rotation accuracy across large and small transformation tasks

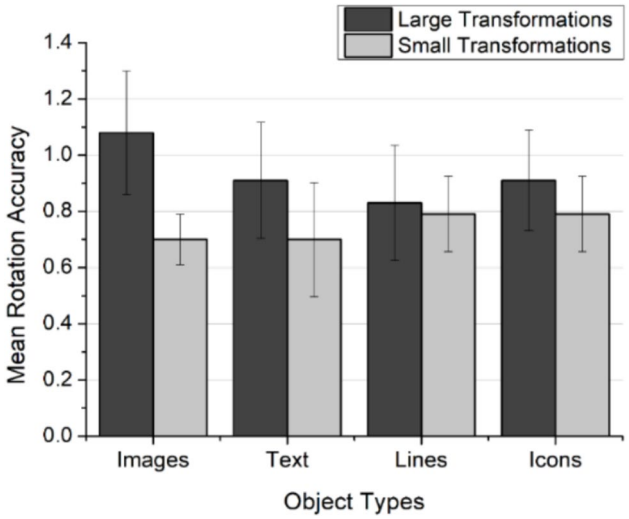

Fig. 5 Mean rotation accuracy across object types

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 (Fig. 4). 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. Figure 5 presents the mean rotation accuracy across all four types of objects.

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 [10].

4.2.4 Speech recognition performance

Overall, a total of 1164 speech commands were issued across all tasks and participants. Figure 4 represents the number of commands across all objects and transformations. Among these, 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.

4.2.5 Qualitative feedback

Overall, participants provided 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*" (P8). 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: "*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" (P10). No participants mentioned any unique challenges associated with rotating different object types. Three participants (P5, P7, P11) suggested that some form of support or visual hints (e.g., guidelines) could reduce the use of speech commands for larger transformations. Three participants (P3, P7, P9) 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*" (P9).

4.3 Study discussion

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 participants also highlighted some clear usability issues (in particular, around estimating transformation angles). To explore this area further, an additional study was conducted to investigate the efficacy of alternative voice-controlled rotation approaches.

5 Study 3—evaluation of voice controlled rotation techniques

This 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.

5.1 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 [42] used for recognizing speech commands. The prototype interface presented a logo design task for a fictional professional designer (Fig. 6).

The interaction object (Fig. 6e) was presented as an element of a logo design which could be rotated using appropriate voice commands in each condition. A target placeholder (Fig. 6f) 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 (Fig. 6g) 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 recognizer before starting each task. Based on the previous exploratory study, the same rotation approach was used again to provide a control condition to compare against two new techniques: "Fixed-Jumps" and "Animation-Rotation".

5.1.1 Baseline-rotation

Baseline-Rotation uses simple commands such as "right 10" and "left 20" that were used previously in the exploratory study. When a user issues a command (e.g., "right 10"), the relevant object is rotated 10 degrees in a clockwise direction. Similarly, when a voice command such as "left 10" is issued, the interaction object rotates 10 degrees in an anti-clockwise direction. No additional commands in this approach support rotation of objects. Figure 7 presents an example of how the objects are rotated using Baseline-Rotation.

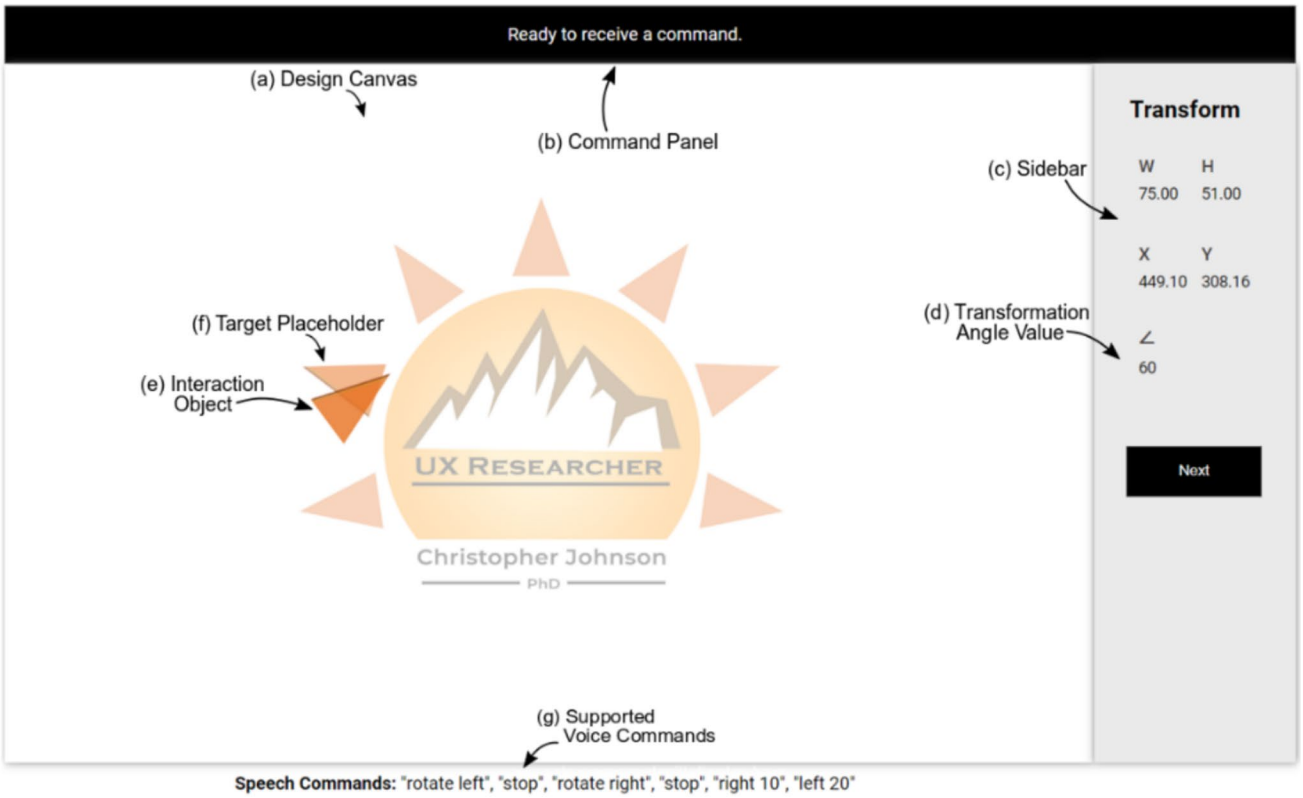

Fig. 6 Main object rotation prototype demonstrating interface elements **a** design canvas, **b** command panel, **c** sidebar, **d** transformation angle value, **e** interaction object, **f** target placeholder, and **g** supported voice commands

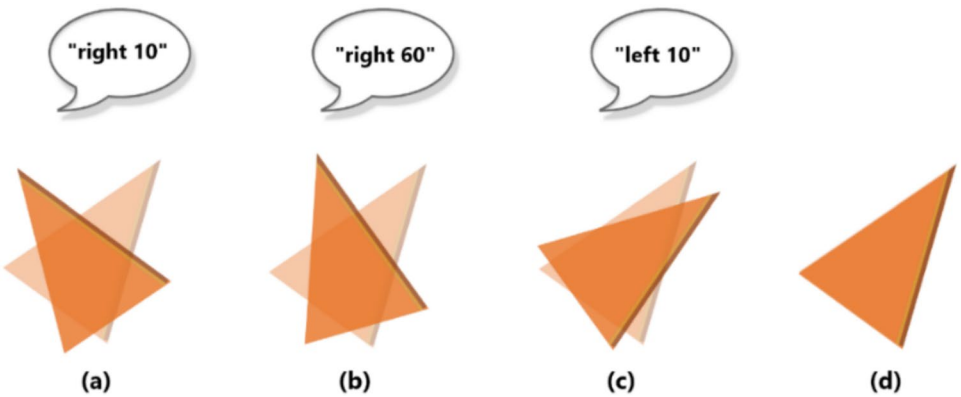

Fig. 7 Baseline-Rotation: **a** the darker object is being rotated in clockwise direction using command "right 10", **b** user issues command "right 60" to take object towards target rotation position, **c** object goes beyond target position, thus user issues another "left 10" command, **d** object is successfully positioned at target rotation angle

5.1.2 Baseline rotation + fixed-jumps

This approach continuously rotates the interaction object via 15 degree "jumps"—to initiate this method, users can issue "rotate left" or "rotate right" commands (Fig. 8). 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. The rationale for focusing on this approach was to investigate whether a similar feature widely available in mainstream creative applications (i.e., Adobe Photoshop, XD, Figma, and Inkscape) could be tailored for voice interaction and whether this presented any interaction benefits. This feature typically involves a user holding the shift key to lock rotation and mouse dragging movements to rotate objects in 15-degree jumps. By tailoring this for speech input, we wanted to explore if this type of controlled and

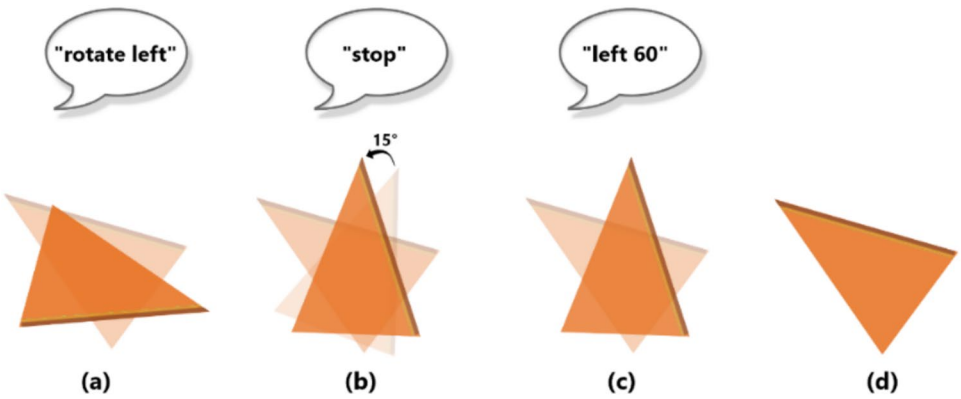

Fig. 8 Fixed-Jumps approach: **a** user issues commands "rotate left" to rotate object in an anticlockwise direction **b** object is rotated continuously in 15 degrees when the user issues a "stop" as the object reaches closer to the target rotation **c** after stopping the rotation, the user issues a "left 60" command, **d** object is placed at target rotation angle

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.

5.1.3 Baseline rotation + animation-rotation

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 object with increased speed, as well as a "slower" command to reduce the rotation speed. A user can also issue a "stop" command to stop the current rotation of an object (Fig. 9). 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. The rationale for developing this approach was to explore whether it can facilitate users in efficiently rotating objects through reducing the number of speech commands that need to be issued (similar to Fixed-Jumps). Furthermore, the option to dynamically alter the rotation speed also presents additional control over Fixed-Jumps that can potentially 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.

5.2 Methodology

5.2.1 Participants

A total of 25 participants with physical impairments (15 male and 10 female) 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).

5.2.2 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 utilize a switch input (a keyboard or assistive tool of their choice) to enable the speech recognizer. Twelve participants used a physical keyboard (i.e., spacebar

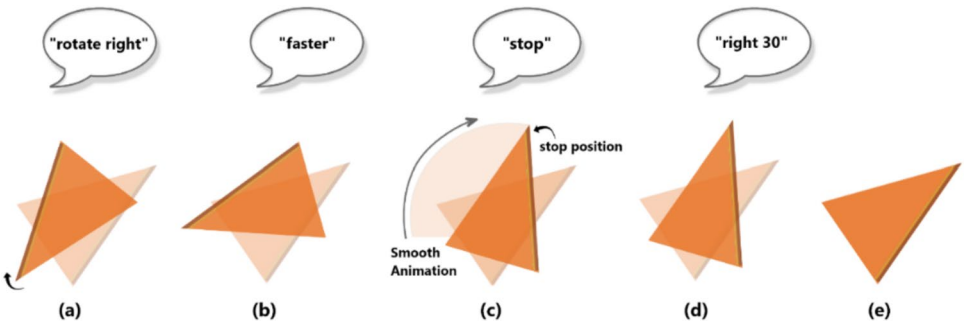

Fig. 9 Animation-Rotation: **a** object is being rotated smoothly (without 15 degrees jumps) in a clockwise direction, **b** user issues "faster" command to increase rotation speed, **c** user stops the rotation close to the target position, **d** user issues command "right 30" to adjust object over target placeholder, **e** object has been adjusted over target rotation position

Table 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	IP: Average; GM: 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 (Since 2017)	Hand tremors; Shooting pain in hands and arms; Pain in wrists; muscle weakness	IP: Average; GM: Expert; ST: Dragon software; AT: NA

Table 5 (continued)

ID	Age/gender	Physical impairments	Condition details	Technical experience
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

key), six used Dragon software [14] via a "press spacebar" command, four utilized a foot pedal, and three used an eye tracker (with an on-screen keyboard). The Google Chrome browser was utilized to ensure compatibility with the Web Speech API [42].

5.2.3 Procedure

Institutional ethical review board approval was initially obtained. Participants were initially 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. The researcher provided pre-test instructions to ensure participants understood the purpose of the study and then redirected them to a consent page, followed by a pre-test survey where questions focused around demographic information, graphic design, prototyping and speech technology experience.

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 min). Participants then selected a button available within 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), and the order of interaction modes and tasks were randomized to minimise 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 (Fig. 10).

Once participants felt they had accurately completed the rotation activity, they clicked the "Next" button 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 participants' perceptions of the different rotation techniques. Testing sessions lasted between 50 to 60 min in total and were video recorded for later analysis.

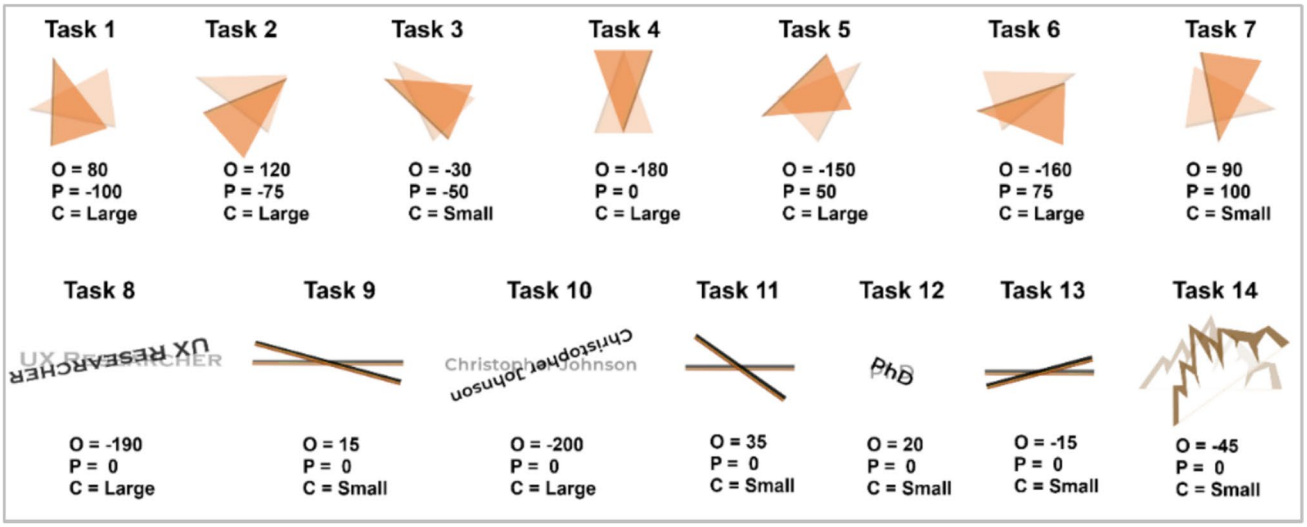

Fig. 10 Interaction Object (O)—in darker shade, Target Placeholder (P)—semi-transparent object located underneath the interaction object, and rotation transformation category (C)—large or small

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 from when a user issued the first rotation command until they then ceased interaction by uttering the last voice command to complete a task. 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 classified into three categories: "Speech Misrecognition", "System Error", and "Unsupported Commands". Moreover, SUS was used to evaluate the perceived usability of each interaction approach.

5.3 Results

Shapiro–Wilk's test [38] 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 utilized to analyze the differences in task completion time, total number of speech commands and SUS scores for each interaction approach. A non-parametric Friedman test [47] with Wilcoxon signed rank was used to analyze the differences in rotation accuracy and speech recognition errors between conditions. A breakdown of key metrics across each condition is provided in Table 6.

5.3.1 Task completion time

The mean task completion time for Baseline-Rotation was 7.80 min (SD = 0.63), 8.81 min (SD = 0.54) for Fixed-Jumps, and 7.48 min (SD = 0.54) for Animation-Rotation. A statistically significant difference was found between the three conditions in terms of task completion time ($F(2, 48) = 35.13$, $p < 0.001$, partial $\eta^2 = 0.98$). Post-hoc tests identified a significant difference between Baseline-Rotation and Animation-Rotation ($p < 0.05$), Baseline-Rotation and Fixed-Jumps ($p < 0.05$), as well as Animation-Rotation and Fixed-Jumps ($p < 0.05$) (Fig. 11).

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-Rotation. 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-Rotation ($Z = -3.38$, $p < 0.001$), and Animation-Rotation 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 12 demonstrates that the mean rotation accuracy of Animation-Rotation is higher as the difference between the final positions of interaction objects and target placeholders is lower compared to the other conditions.

Table 6 Object transformation tasks, task completion time (seconds) and number of speech commands (Com. (*n*)) across the three speech interaction techniques

Tasks (Transformations)	Baseline-rotation		Fixed-jumps		Animation	
	Time (Sec)	Com. (*n*)	Time (Sec)	Com. (*n*)	Time (Sec)	Com. (*n*)
Task 1 (Large)	M = 40.96 SD = 0.11	N = 232 M = 9.28 SD = 1.86	M = 45.43 SD = 0.12	N = 258 M = 10.32 SD = 1.61	M = 39.87 SD = 0.12	N = 225 M = 9.00 SD = 1.26
Task 2 (Large)	M = 40.19 SD = 0.18	N = 224 M = 8.96 SD = 1.51	M = 44 SD = 0.13	N = 248 M = 9.92 SD = 1.23	M = 37.14 SD = 0.12	N = 214 M = 8.56 SD = 1.06
Task 3 (Small)	M = 22.88 SD = 0.22	N = 146 M = 5.84 SD = 0.92	M = 28.69 SD = 0.21	N = 159 M = 6.36 SD = 1.29	M = 21.76 SD = 0.13	N = 141 M = 5.64 SD = 0.62
Task 4 (Large)	M = 38.23 SD = 0.10	N = 217 M = 8.68 SD = 1.56	M = 45.19 SD = 0.08	N = 248 M = 9.92 SD = 1.46	M = 37.17 SD = 0.10	N = 212 M = 8.48 SD = 1.33
Task 5 (Large)	M = 41.35 SD = 0.13	N = 245 M = 9.80 SD = 1.92	M = 45.96 SD = 0.11	N = 266 M = 10.64 SD = 1.78	M = 39.89 SD = 0.11	N = 219 M = 8.76 SD = 1.58
Task 6 (Large)	M = 39.33 SD = 0.12	N = 233 M = 9.32 SD = 1.91	M = 47.73 SD = 0.08	N = 252 M = 10.08 SD = 1.62	M = 38.30 SD = 0.10	N = 214 M = 8.56 SD = 1.29
Task 7 (Small)	M = 27.57 SD = 0.13	N = 152 M = 6.08 SD = 1.13	M = 29.49 SD = 0.10	N = 173 M = 6.92 SD = 1.05	M = 27.50 SD = 0.12	N = 147 M = 5.88 SD = 0.86
Task 8 (Large)	M = 39.71 SD = 0.09	N = 239 M = 9.56 SD = 2.04	M = 43.79 SD = 0.08	N = 254 M = 10.16 SD = 1.93	M = 37.93 SD = 0.08	N = 221 M = 8.84 SD = 1.59
Task 9 (Small)	M = 27.15 SD = 0.11	N = 153 M = 6.12 SD = 1.10	M = 30.91 SD = 0.07	N = 174 M = 6.96 SD = 0.99	M = 25.63 SD = 0.17	N = 147 M = 5.88 SD = 0.81
Task 10 (Large)	M = 42.74 SD = 0.13	N = 247 M = 9.88 SD = 1.96	M = 45.57 SD = 0.11	N = 263 M = 10.52 SD = 1.57	M = 40.35 SD = 0.08	N = 242 M = 9.68 SD = 1.22
Task 11 (Small)	M = 29.44 SD = 0.16	N = 165 M = 6.60 SD = 0.94	M = 32.32 SD = 0.15	N = 184 M = 7.36 SD = 0.97	M = 28.52 SD = 0.13	N = 157 M = 6.28 SD = 1.21
Task 12 (Small)	M = 27.68 SD = 0.13	N = 158 M = 6.32 SD = 0.67	M = 30.56 SD = 0.09	N = 169 M = 6.76 SD = 0.91	M = 26.31 SD = 0.08	N = 147 M = 5.88 SD = 0.76
Task 13 (Small)	M = 26.29 SD = 0.14	N = 149 M = 5.96 SD = 0.96	M = 29.17 SD = 0.16	N = 157 M = 6.28 SD = 1.11	M = 25.23 SD = 0.08	N = 136 M = 5.44 SD = 0.63
Task 14 (Small)	M = 24.51 SD = 0.15	N = 144 M = 5.76 SD = 0.90	M = 29.79 SD = 0.09	N = 163 M = 6.52 SD = 1.17	M = 23.36 SD = 0.10	N = 131 M = 5.24 SD = 0.99

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-Rotation (Fig. 13). 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-Rotation ($p < 0.05$), Baseline-Rotation and Fixed-Jumps ($p < 0.05$), and between Animation-Rotation 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-Rotation. 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

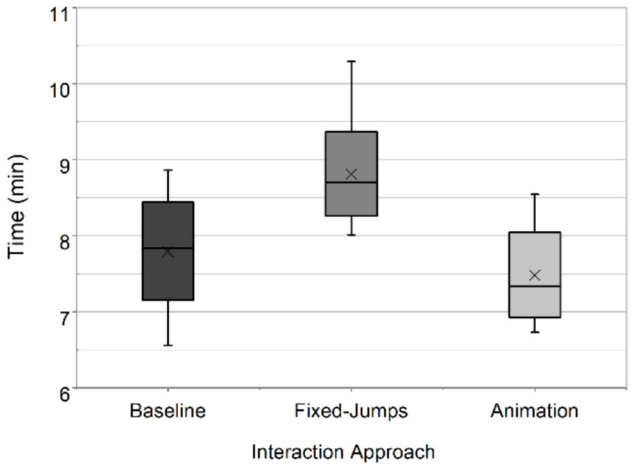

Fig. 11 Mean task completion time across the three speech-controlled object rotation approaches

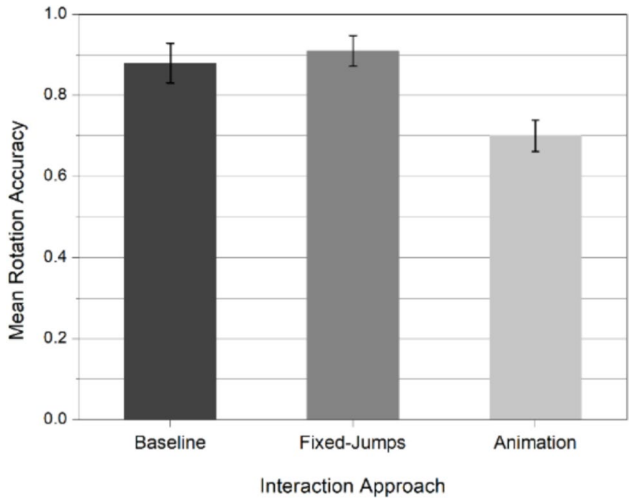

Fig. 12 Mean rotation accuracy across all three object rotation approaches

52 (2.04%) related to Animation-Rotation. 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-Rotation. 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 $p > 0.05$).

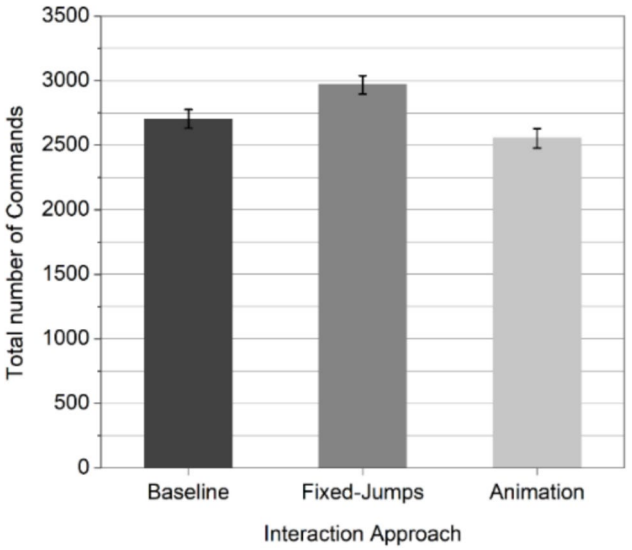

Fig. 13 Total number of speech commands

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-Rotation. The Baseline-Rotation and Animation-Rotation scores can be labelled as "Good" while Fixed-Jumps Rotation can be labelled as "Above Average" [10]. 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 also displayed a significant difference between Baseline-Rotation and Animation-Rotation ($p < 0.05$), Baseline-Rotation and Fixed-Jumps ($p < 0.05$), and between Animation-Rotation and Fixed-Jumps ($p < 0.05$) (Fig. 14).

5.3.5 Qualitative feedback

Fifteen participants preferred Animation-Rotation, eight highlighted a preference for Baseline-Rotation, whilst two participants preferred Fixed-Jumps. Participants who provided positive feedback regarding Animation-Rotation suggested it was both "*faster*" and more "*effective*" than Baseline-Rotation and Fixed-Jumps. Participants also highlighted that the Animation-Rotation approach enabled them to efficiently rotate objects for larger transformations: *"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"* (P14). Eight participants who preferred Baseline-Rotation commented that they found this approach "*easy to use*", "*straightforward*" and having "*less complexity*"

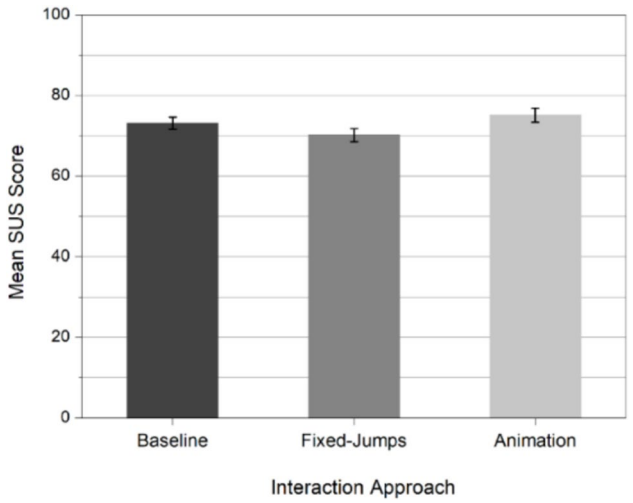

Fig. 14 The mean SUS score across the three rotation interaction approaches

than other approaches: *"…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"* (P9). 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"* (P6). 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"* (P13).

6 Discussion

This paper has explored new speech-controlled object rotation techniques for manipulating graphical objects based within a digital canvas. An initial elicitation study identified the types of commands that disabled 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 voice commands identified. An exploratory study utilizing this prototype demonstrated that people with physical impairments could successfully manipulate the rotation of digital assets, although some key usability challenges were highlighted (i.e., estimating transformation angles). To address these challenges, an updated version of the prototype was developed and evaluated with disabled users. Results found that the Animation-Rotation approach was faster, more accurate, and usable than the other two approaches, as well as demonstrating consistent trends across individual tasks for all large and small transformations. Subjective feedback from participants also highlighted a preference for the Animation-Rotation technique over Baseline-Rotation and Fixed-Jumps. This research therefore contributes new knowledge and insights around how people with physical impairments can effectively and efficiently rotate digital assets via voice interaction.

One limitation of the research is that the task was focused on a specific scenario associated with a 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. 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-Rotation approach and had to monitor it closely to choose when to issue the "stop" command (or when it was appropriate to use slower or 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-Rotation in the context of the tasks completed during the study. 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 placeholder. 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 potential limitation of the presented speech interaction approaches is social acceptability as users are always required to speak aloud which might be socially inconvenient or uncomfortable within public spaces. Hence the voice-controlled methods developed are likely to be more appropriate for use within private spaces or when speech interaction is a user's preference [31, 32]. Finally, whilst we explored three different speech-controlled object rotation approaches, it should be noted that there are additional approaches that could also be investigated which may present more efficient and intuitive methods. For example, it was clear from the studies conducted that some participants

found it challenging on occasions to estimate correct transformation angles. An alternative approach could utilize a clock interface around an object where users can state a number associated with the long hand of clock (e.g., "right 10" to rotate an object to 10 past the hour). The use of non-speech verbal sounds such as humming and whistling [39] or vowel sounds [17] also represent alternative approaches for rotating objects at different rotation angles, although further research is required to understand the efficacy of these methods.

While the speech interaction approaches presented for object rotation have been designed primarily for people with physical impairments, these methods also hold potential to support non-disabled designers. For instance, voice-controlled rotation approaches can augment a designer's creative workflow when using a mouse, keyboard or a stylus for creative activities thus facilitating more efficient rotation transformations (e.g., through enabling users to initiate rotation 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 [26]. Furthermore, the use of voice control within collaborative design activities with colleagues could also enable non-disabled users to seamlessly communicate verbally with others alongside controlling objects via speech input (which may present interaction benefits), although further work is required to fully understand the challenges and opportunities associated with this approach.

Similarly, whilst this study presented and focused on manipulating the rotation of graphical objects to support creative design work, 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 [13, 49], simultaneous transformation of multiple objects, and arranging interface elements to enhance the appearance of digital designs [46], as well as manipulating 3D objects [6, 43, 44]. The study also only focused on four different types of common assets (i.e., images, text, lines, and icons)—whilst this presented insights around users rotation experiences across object types, it will also be important in future work to explore a wider range of assets such as custom shapes, multi-line text objects, and objects of varying sizes, [11, 23, 24, 35, 48], as well as how users perceive these approaches for other manipulations such as positioning and resizing objects in different design scenarios. The use of natural language commands for rotating and manipulating objects via speech commands such as "*rotate 100 degrees in a clockwise direction*" or "*rotate 20 degrees to the left*" is also an underexplored area that may present some interaction benefits. However, this may also potentially result in user frustration if current natural language models are unable to reliably and consistently determine a user's intention (thus resulting in incorrect actions being performed by the system) [37].

7 Conclusion

This research presents the first empirical evaluations to explore different voice-controlled 2D object rotation techniques (Baseline-Rotation, Fixed-Jumps, and Animation-Rotation) for supporting people with physical impairments. Results found that the Animation-Rotation approach was more efficient and usable than the other two methods—subjective feedback also supported these findings with participants sharing positive perceptions around the usability of this approach. Feedback from disabled participants across all three evaluations has also highlighted useful and fruitful insights for future research which can help to inform the design of more inclusive creative design environments for people with physical impairments.

Author contributions All authors contributed to the preparation of the paper. Material preparation, data collection and analysis were performed by Farkhandah Aziz and Chris Creed. The first draft of the manuscript was written by Farkhandah Aziz, and all authors Chris Creed, Maite Frutos and Ian Williams provided feedback on previous versions of the paper. All authors read and approved the final manuscript prior to submission.

Funding The authors did not receive support from any organization for the submitted work.

Data availability All data generated or analyzed during this study are included in this published article.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval The research work received ethical approval from Research Ethics Committee at Birmingham City University, UK (Reference: Komal /#10132 /sub2 /R(B) /2022 /Mar /CEBE FAEC).

Informed consent Informed consent was received from all participants prior to taking part in the study.

Consent for publication All authors consent to the publication of this article.

References

1. Adobe. 2023. Photoshop apps—desktop, mobile, and tablet | Photoshop.com. Retrieved Mar 17, 2023 from https://www.adobe.com/products/photoshop.html
2. Adobe Inc. 2023. Adobe Illustrator CS6: industry-leading vector graphics software. Retrieved Mar 17, 2023 from https://www.adobe.com/uk/products/illustrator.html
3. Adobe Inc. 2023. Adobe XD | Fast & Powerful UI/UX Design & Collaboration Tool. Retrieved Mar 17, 2023 from https://www.adobe.com/uk/products/xd.html
4. Figma. 2023. Figma: The collaborative interface design tool. Retrieved Mar 17, (2023). From https://www.figma.com
5. Nuance Communications. 2023. Dragon Speech Recognition—Get More Done by Voice | Nuance. Retrived Mar 17, (2023). From https://www.nuance.com/dragon.html
6. Alibay, F., Kavakli, M., Chardonnet, J.R. and Baig, M.Z.: The usability of speech and/or gestures in multi-modal interface systems. In: Proceedings of the 9th international conference on computer and automation engineering. (2017). https://doi.org/10.1145/3057039.3057089
7. Alsuraihi, M.M., Rigas, D.I.: 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*. 1–4. Retrieved Mar 17, 2023 (2007). From https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2007.42
8. Aziz, F., Creed, C., Sarcar, S., Frutos-Pascual, M., Williams, I.: Voice snapping: inclusive speech interaction techniques for creative object manipulation. In: designing interactive systems conference. 1486–1496. (2022). https://doi.org/10.1145/3532106.3533452
9. Aziz, F., Creed, C., Sarcar, S., Frutos-Pascual, M., Williams, I.: Inclusive voice interaction techniques for creative object positioning. In: proceedings of the 2021 international conference on multimodal interaction. 461–469, 2021. https://doi.org/10.1145/3462244.3479937
10. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Stud. **4**(3), 114–123 (2009). https://doi.org/10.5555/2835587.2835589
11. Creed, C., Frutos-Pascual, M., Williams, I.: Multimodal gaze interaction for creative design. In: proceedings of the 2020 CHI conference on human factors in computing systems. 1–13, (2020).https://doi.org/10.1145/3313831.3376196
12. Dai, L., Goldman, R., Sears, A., Lozier, J.: Speech-based cursor control: a study of grid-based solutions. ACM SIGACCESS Access. Comput. **77–78**, 94–101 (2003). https://doi.org/10.1145/1029014.1028648
13. Dai, L., Goldman, R., Sears, A., Lozier, J.: Speech-based cursor control using grids: modelling performance and comparisons with other solutions. Behav. Inf. Technol. **24**(3), 219–230 (2005). https://doi.org/10.1080/01449290412331328563
14. Dragon. (2023). Nuance Communications. Dragon Speech Recognition—Get More Done by Voice | Nuance. Retrieved Mar 17, 2023, from https://www.nuance.com/dragon.html
15. Elepfandt, M., Grund, M.: 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. 1–3 (2012). https://doi.org/10.1145/2401836.2401848
16. Gourdol, A.P., Nigay, L., Salber, D., Coutaz, J.: Two case studies of software architecture for multimodal interactive systems: voicepaint and a voice-enabled graphical notebook. Eng. Hum. Comput. Interact. **92**, 271–84 (1992)
17. Harada, S., Saponas, T.S., Landay, J.A.: VoicePen: augmenting pen input with simultaneous non-linguistic vocalization. In: proceedings of the 9th international conference on multimodal interfaces, ICMI'07. 178–185 (2007). https://doi.org/10.1145/1322192.1322225
18. Harada, S., Wobbrock, J.O., Landay, J.A.: VoiceDraw: a hands-free voice-driven drawing application for people with motor impairments. In ASSETS'07: proceedings of the ninth international ACM SIGACCESS conference on computers and accessibility 27–34 (2007.). https://doi.org/10.1145/1296843.1296850
19. Harada, S., Wobbrock, J.O., Malkin, J., Bilmes, J.A., Landay, J.A.: 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) (2009). https://doi.org/10.1145/1518701.1518757
20. Hauptmann, A.G.: Speech and gestures for graphic image manipulation. In: Proceedings of the SIGCHI conference on Human factors in computing systems 241–245 (1989). https://doi.org/10.1145/67449.67496
21. Hiyoshi, M., Shimazu, H.: Drawing pictures with natural language and direct manipulation. In: COLING 1994 volume 2: the 15th international conference on computational linguistics (1994). https://doi.org/10.3115/991250.991262
22. House, B., Malkin, J., Bilmes, J.: The VoiceBot: a voice controlled robot arm. In: proceedings of the SIGCHI conference on human factors in computing systems. 183–192 (2009). https://doi.org/10.1145/1518701.1518731
23. Hu, R., Zhu, S., Feng, J., Sears, A.: Use of speech technology in real life environment. In: universal access in human-computer interaction. Applications and services: 6th international conference, UAHCI 2011, held as part of HCI international 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part IV 6. 62-71 (2011). https://doi.org/10.1007/978-3-642-21657-2_7
24. Kamel, H.M., Landay, J.A.: A study of blind drawing practice: creating graphical information without the visual channel. In: proceedings of the fourth international ACM conference on Assistive technologies. 34–41 (2000). https://doi.org/10.1145/354324.354334
25. Karimullah, A.S., Sears, A.: Speech-based cursor control. In: Proceedings of the fifth international ACM conference on Assistive technologies. 178–185 (2002). https://doi.org/10.1145/638249.638282
26. Kim, Y.S., Dontcheva, M., Adar, E., Hullman, J.: Vocal shortcuts for creative experts. In: proceedings of the 2019 CHI conference on human factors in computing systems. 1–14 (2019). https://doi.org/10.1145/3290605.3300562
27. Laput, G.P., Dontcheva, M., Wilensky, G., Chang, W., Agarwala, A., Linder, J., Adar, E.: Pixeltone: a multimodal interface for image editing. In: proceedings of the SIGCHI conference on human factors in computing systems. 2185–2194 (2013). https://doi.org/10.1145/2470654.2481301
28. Laviola, J.J., Katzourin, M.: An exploration of non-isomorphic 3D rotation in surround screen virtual environments. In: 2007 IEEE symposium on 3D user interfaces. IEEE (2007). https://doi.org/10.1109/3DUI.2007.340774
29. Milota, A.D. : Modality fusion for graphic design applications. In: proceedings of the 6th international conference on multimodal interfaces. 167–174 (2004). https://doi.org/10.1145/1027933.1027963
30. Nishimoto, T., Shida, N., Koayashi, T., Shirai, K.: Improving human interface drawing tool using speech, mouse and key-board.

In: proceedings 4th ieee international workshop on robot and human communication. 107–112 (1995). IEEE. https://doi.org/10.1109/ROMAN.1995.531944
31. O'Shaughnessy, D.: Automatic speech recognition: history, methods and challenges. Pattern Recogn. **41**(10), 2965–2979 (2008). https://doi.org/10.1016/j.patcog.2008.05.008
32. Oviatt, S., Cohen, P., Wu, L., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D.: Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Hum. Comput. Interact. **15**(4), 263–322 (2000). https://doi.org/10.1207/S15327051HCI1504_1
33. Pausch, R., Leatherby, J.H.: An empirical study: adding voice input to a graphical editor. In: Journal of the american voice input/output society (1991). Retrieved Mar 17, 2023 , from https://citeseerx.ist.psu.edu/doc_view/pid/217e9d8ccf975a99e0c910e2ed12a3d512154c8b
34. Poupyrev, I., Weghorst, S., Fels, S.: Non-isomorphic 3D rotational techniques. In: Proceedings of the SIGCHI conference on Human factors in computing systems. 540–547 (2000). https://doi.org/10.1145/332040.332497
35. Schaadhardt, A., Hiniker, A., Wobbrock, J.O.: Understanding blind screen-reader users' experiences of digital artboards. In: Proceedings of the 2021 CHI conference on human factors in computing systems. 1–19 (2021). https://doi.org/10.1145/3411764.3445242
36. Sedivy, J., Johnson, H.: Supporting creative work tasks: the potential of multimodal tools to support sketching. In: proceedings of the 3rd conference on creativity & cognition. 42–49 (1999). https://doi.org/10.1145/317561.317571
37. Serai, P., Sunder, V., Fosler-Lussier, E.: Hallucination of speech recognition errors with sequence to sequence learning. IEEE/ACM Trans. Audio, Speech, Lan-Guage Process. **30**, 890–900 (2022). https://doi.org/10.1109/TASLP.2022.3145313
38. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (Complete Samples). Biometrika. 591–611 (1965). https://doi.org/10.2307/2333709
39. Sporka, A.J., Kurniawan, S.H., Mahmud, M., Slavík, P.: 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 213–220 (2006). https://doi.org/10.1145/1168987.1169023
40. Srinivasan, A., Dontcheva, M., Adar, E., Walker, S.: Discovering natural language commands in multimodal interfaces. In: proceedings of the 24th international conference on intelligent user interfaces. 661–672 (2019). https://doi.org/10.1145/3301275.3302292
41. Van der Kamp, J., Sundstedt, V.: Gaze and voice controlled drawing. In: Proceedings of the 1st conference on novel gaze-controlled applications. 1–8 (2011).https://doi.org/10.1145/1983302.1983311
42. Web Speech API: MND | Developer.mozilla.org—Web APIs. Retrieved Mar 17, (2023). from https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API
43. Williams, A.S., Garcia, J., Ortega, F.: Understanding multimodal user gesture and speech behavior for object manipulation in augmented reality using elicitation. IEEE Trans. Visual Comput. Graphics **26**(12), 3479–3489 (2020). https://doi.org/10.1109/TVCG.2020.3023566
44. Williams, A.S., Ortega, F.R.: Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation. In: Proceedings of the ACM on human-computer interaction, 4(ISS), 1–21 (2020).https://doi.org/10.1145/3427330
45. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI conference on human factors in computing systems. 1083–1092 (2009). https://doi.org/10.1145/1518701.1518866
46. Xu, P., Fu, H., Igarashi, T., Tai, C.L.: Global beautification of layouts with interactive ambiguity resolution. In: Proceedings of the 27th annual ACM symposium on User interface software and technology. 243–252 (2014). https://doi.org/10.1145/2642918.2647398
47. Zimmerman, D.W., Zumbo, B.D.: Relative power of the wilcoxon test, the friedman test, and repeated-measures ANOVA on ranks. J. Exp. Educ. **62**(1), 75–86 (1993). https://doi.org/10.1080/00220973.1993.9943832
48. Zhu, S., Ma, Y., Feng, J., Sears, A.: Speech-based navigation: improving grid-based solutions. In: human-computer interaction–INTERACT 2009: 12th IFIP TC 13 international conference, Uppsala, Sweden, August 24–28, 2009, Proceedings, Part I 12. 50–62. Springer Berlin Heidelberg (2009). https://doi.org/10.1007/978-3-642-03655-2_6
49. Zhu, S., Feng, J., Sears, A.: Investigating grid-based navigation: the impact of physical disability. ACM Trans. Access. Comput. (TACCESS) **3**(1), 1–30 (2010). https://doi.org/10.1145/1838562.1838565