Creative Flow in VR Sketching: Comparing Unimodal and Multimodal Input Techniques in the Creative Process

Richard Rodriguez*
Computer Science
Colorado State University
Fort Collins, Colorado, USA
richir@colostate.edu

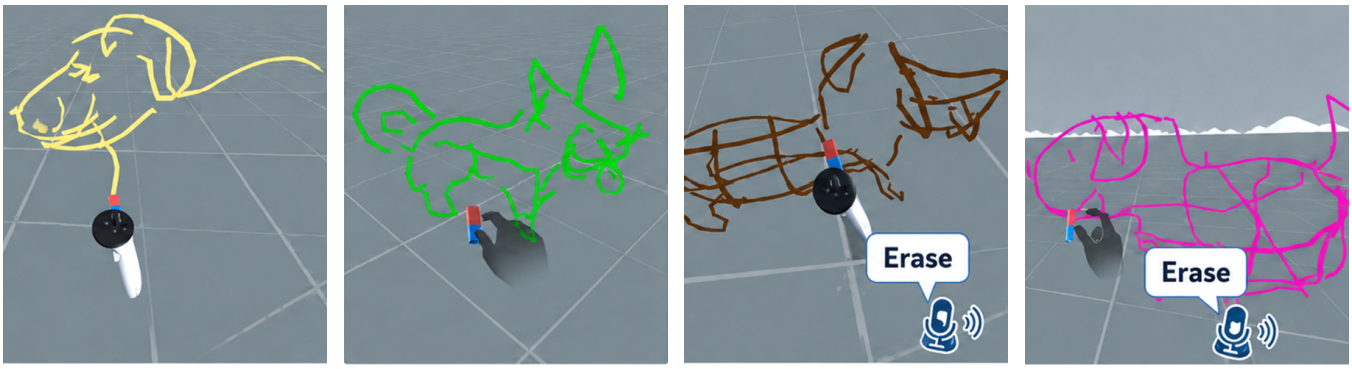
Zahra Borhani*
Computer Science
Colorado State University
Fort Collins, Colorado, USA
zahra.borhani@colostate.edu

Tanmay Naidu
Computer Science
Colorado State University
Fort Collins, Colorado, USA
tanmay.naidu@colostate.edu

Anil Ufuk Batmaz
Department of Computer Science & Software Engineering
Concordia University
Montreal, Quebec, Canada
ufuk.batmaz@concordia.ca

Mayra Donaji Barrera Machuca
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
mbarrera@ucalgary.ca

Francisco Raul Ortega
Computer Science
Colorado State University
Ft. Collins, Colorado, USA
fortega@colostate.edu



(a) *Controller* (b) *Microgesture* (c) *Controller+Speech* (d) *Microgesture+Speech*

Figure 1: sketches created by a participant under the four experimental interaction conditions: Unimodal (*Controller, Microgesture*) and Multimodal (*Controller+Speech, Microgesture+Speech*) for an erasing action.

Abstract

3D sketching allows users to create directly in 3D space with sketching tools, menus, and commands embedded within the virtual environment. However, creative work can be disrupted when artists must pause to locate menus, remember button mappings, or switch tools. Multimodal input, particularly speech paired with manual input, may reduce these interruptions by enabling rapid access to commands while users continue sketching. This paper examines how base input modality and speech augmentation shape creative flow during immersive 3D sketching. We conducted a 2 x 2 within-subject study with 24 participants, comparing controller and microgesture input, each with and without speech commands. Results showed that speech augmentation significantly increased flow, usability, and presence, but the size of these improvements was modest. *Microgesture+Speech* resulted in the highest descriptive flow scores. However, the results did not show a significant difference, indicating that microgesture input did not benefit more from speech than controller input. These findings suggest that multimodal systems for 3D sketching using speech can support more fluid creative workflows while retaining the benefits of conventional inputs.

CCS Concepts

• **Human-centered computing** → **User studies**; **Usability testing**; **Gestural input**; **User centered design**.

Keywords

Virtual Reality Sketching, 3D sketching, Multimodal Interaction, Microgesture Input, Creative Flow, Controller input, User study, Usability

ACM Reference Format:
Richard Rodriguez, Zahra Borhani, Tanmay Naidu, Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, and Francisco Raul Ortega. 2026. Creative Flow in VR Sketching: Comparing Unimodal and Multimodal Input Techniques in the Creative Process. In *32nd ACM Symposium on Virtual Reality Software*

*These authors contributed equally to this work.

This work is licensed under a Creative Commons Attribution 4.0 International License.

VRST '26, Sendai, Japan

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2811-2/2026/11
https://doi.org/10.1145/3822517.3848721

and Technology (VRST '26), November 16–18, 2026, Sendai, Japan. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3822517.3848721

1 Introduction

In recent years, Virtual Reality (VR) has emerged as a promising platform for creative work. For example, VR app stores currently offer several applications that allow users to sketch, paint, and draw in a 3-dimensional (3D) environment, where they can walk freely to sketch at different points in space, or verify their work from different angles [2]. This creates a new avenue for everyone to express their creativity and explore what they are capable of. Applications such as Open Brush [29], Gravity Sketch [23], and Quill [53] have demonstrated that they are capable of setting free the inner artists, even when one is not an expert in sketching [39]. With steady updates, these apps add new features, allowing them to evolve and, in turn, allowing users to continue exploring.

In creative VR applications such as games [7] and VR sketching [27], flow is often considered a key indicator of optimal user experience [7]. Flow captures the extent to which a person becomes deeply absorbed in a task while maintaining a sense of control and focus [14, 45]. In creative work, flow is especially important because the tool should support the user's attention to the emerging idea rather than interrupting it. When artists sketch, interruptions such as searching through menus or switching tools can shift attention away from the creative process and toward system operation [24, 33].

Creative flow is influenced by several factors identified in prior research, including the challenge-skill balance, action-awareness merging, unambiguous feedback, and a sense of control [25, 48]. In VR sketching, these dimensions are closely tied to the input technique used by the artist, as different input modalities can provide different levels of support, constraint, precision, and cognitive demand [49]. For example, controller-based input may provide precision and familiarity, but it may also require users to navigate menus or remember button mappings, which can disrupt concentration. Conversely, a more embodied input technique, such as gesture-based or microgesture-based input, may feel more natural and immersive, but can introduce uncertainty if recognition is unreliable or if users accidentally trigger actions.

To examine how the base input modality may influence creative flow in VR sketching, we compared controller-based input with a hand-based interaction approach that combines midair gestures and microgestures. Microgestures are a subset of gestures characterized by small-scale, fine-grained, low-effort movements, typically subtle finger or thumb motions used to perform short interaction tasks [37, 38]. While midair gestures generally involve larger arm or hand movements, microgestures are better suited to situations in which large movements are inconvenient, tiring, or disruptive [13, 26]. A pinch, for example, is a microgesture when the arm is stationary and a midair gesture when the arm is in motion; accordingly, we used the pinch as a microgesture for menu interaction and mode switching, and as a midair gesture for drawing, erasing, and selecting strokes. In VR, microgestures can reduce physical fatigue by minimizing the large, repetitive arm movements associated with conventional VR interaction [36, 41], and limiting midair gestures also reduces the frequency with which the artist must aim at virtual menus. This action can interrupt the creative process and degrade accuracy. Our design reserved microgestures for menu interaction and midair gestures for sketching (e.g., drawing). We additionally employed hand poses, in which specific fingers and the wrist are bent into a shape the system recognizes as an interaction; these were paired with midair gestures to enrich certain actions, such as grabbing a stroke by forming a fist and moving it to a new position. In total, our study used six microgestures: a pinch, a thumb tap, and four directional thumb swipes (up, down, left, and right), mapped across the dominant and non-dominant hand, and between drawing and manipulation modes to support common sketching actions, including drawing, selection and manipulation, stroke-size adjustment, undo/redo, erasing, deletion, and menu access.

A second important design dimension is whether the VR sketching experience is affected by the use of a unimodal vs. a multimodal approach. Prior work has shown that combining modalities, such as gesture and speech, often feels more natural to users; however, multimodal interaction can also introduce coordination effort, learning requirements, and the cost of switching between modalities. For example, in VR sketching, speech augmentation may allow users to issue commands without interrupting manual drawing, potentially reducing menu navigation and helping them stay focused on the sketch. At the same time, speech can introduce cognitive overhead through command recall, coordination effort, or modality switching [59]. Therefore, both the type of input modality and the degree of modality integration, from unimodal to multimodal interaction, can influence users' flow experience in VR.

Despite this connection, most prior studies on VR sketching have evaluated interaction techniques primarily in terms of performance, efficiency, and user preference [5], with less attention to how different interaction approaches affect deeper experiential states such as flow. Similarly, many multimodal interaction studies examine structured activities such as object manipulation or collaborative communication rather than the creative process, which carries different cognitive demands on users. This gap is especially important when considering differences in user expertise and the tradeoff between creative flexibility and ease of interaction.

The motivation behind this work is to examine creative flow in VR sketching through two interaction dimensions. The first dimension is the *base input modality*: controller-based input versus microgesture-based input. The second dimension is *speech augmentation*: unimodal interaction, where users rely only on the base input modality, versus multimodal interaction, where speech is added to support command-based actions. Crossing these two dimensions produces four conditions: *Controller*, *Controller+Speech*, *Microgesture*, and *Microgesture+Speech*. This design allows us to investigate whether creative flow is shaped by the base input modality, whether speech augmentation improves the creative experience, and whether speech affects controller-based and microgesture-based interaction differently.

To evaluate these effects, we conducted a within-subjects VR sketching study in which participants completed the same 3D sketching task under all four interaction conditions. We measured creative flow, usability, and presence using the Flow State Scale questionnaire [52], VR Presence questionnaire, and the System Usability Scale (SUS) questionnaire (questionnaires are provided in the supplemental materials). We also measured task completion

time and activation counts during sketching, and collected qualitative feedback via semi-structured interviews at the end of the experiment.

This paper makes three primary contributions. First, it compares controller and microgesture input techniques for creative VR sketching within a 2×2 factorial design that also examines speech augmentation. Second, it evaluates these interaction approaches through the lens of creative flow, together with usability, presence, task completion time, and qualitative feedback. Third, it identifies how speech augmentation influences controller and microgesture-based interaction, offering design insights for future VR sketching systems that aim to support more fluid, natural, and engaging creative workflows.

2 Related Work

2.1 3D Sketching

One of the earliest works in 3D sketching dates back to the 1960s with an application that revolutionized the world. In 1963, Ivan E. Sutherland designed Sketchpad [40, 54], a graphic system that allowed users to create and manipulate shapes in a computer using a light pen, paving the way for future computer-aided drafting (CAD) applications, among other applications. In 2001, Keefe et al. demonstrated via CavePainting [32] that virtual reality (VR) could also serve as an artistic medium alongside the well-known paper-and-pencil. Arora et al. [10] defined 3D sketching as the use of technology to draw in a 3D environment, where an immersive display is used to interact and display the drawing.

VR sketching applications include Brush [29] (an open-source software package that branched from the private app Tilt Brush [22]), Gravity Sketch [23], Quill [53], VR Sketch [4], and Gesture VR [34]. These VR sketching applications have been used by artists and collaborators. For example, Anna Zhilyaeva created the famous *Liberty* [63] VR artwork live in front of an audience.

2.2 Flow Theory

Flow theory was first introduced in 1990 by Mihaly Csikszentmihalyi [6]. Flow investigates the level of immersion that an individual exhibits while performing a specific task. The flow state scale (FSS), developed by Jackson et al. [30] in 1996 is a psychometric tool used to quantify an individual's flow. The FSS addresses the nine dimensions, previously outlined by Csikszentmihalyi. In flow theory, there are nine (9) dimensions in flow, which are shown in Table 5.

Creative flow has been used in the traditional visual arts, such as drawing, painting, and sketching. It has been used to explain the changes a user undergoes during deep concentration on a visual task and how sustained their flow was [18].

Flow can be applied to multiple disciplines. For example, flow theory was used by Madlenovic et al. [43] to find the appropriate combination of tasks to increase the performance of Brain-computer interfaces (BCI) users. BCIs are devices that allow people with physical disabilities to control external devices or perform functions they would otherwise be unable to do [43]. Another field where flow has been used to measure immersion is in games. For example, Sweetser et al. [56] used flow theory to measure the levels of immersion, goals, and skills exhibited by users while playing single-player video games. In the follow-up study, Sweetser et al. [55] further evaluated players, this time in a multiplayer environment, and suggested improvements for future measurements.

Flow theory has also made its way into the realm of VR. Huang et al. [28] measured the dimensions of flow in *La Camera Insabiatta*, which won the Best VR Experience Award at the 74th Venice International Film Festival [28]. The study consisted of eight VR rooms where participants could appreciate and explore visual art and artistic music, all created by the authors. While the participants were able to appreciate the art, the only movements they were able to accomplish were translations across the various rooms presented. Ruvimova et al. [50] used flow theory to explore ways that VR can be used in an open office environment to improve a participant's flow. Employees in an open office environment can be subject to interference from the outside, which can distract them when trying to get into the workflow. The researchers used VR to mitigate some of these distractions, which improve the employee's workflow.

2.3 Interaction

The use of multisensory perception in VR can improve performance by enhancing the senses [42]. If what the user perceives is important, then multimodal interactions are important in VR, as they allow users to interact with VR systems in the preferred mode for specific tasks. Unimodal interactions, such as the use of controllers, gestures, microgestures, and speech, or multimodal interactions, when two or more unimodal interactions are combined, can aid the user in interacting with the system in an effective way, as it allows the user to interact in the same way that they would in an everyday use case.

2.3.1 Unimodal Interactions. Unimodal interactions refer to using a single input modality. Some unimodal input types include virtual controllers, hand gestures, hand microgestures, pen, speech, gaze, and touch. Multimodal interaction, therefore, refers to the combination of more than one of the aforementioned input modalities. Multimodal interactions were initially demonstrated by Bolt et al. [9]. In their paper, the researchers used gesture and speech as unimodal inputs and their combination as the multimodal input. It has been used in various studies using gesture elicitation, such as William et al. [60], which showed that the combination of gesture and speech is essential in everyday augmented reality (AR). Zhou et al. [64] conducted a study that investigated the use of multimodal interaction in a multi-object environment. The authors found that the combination of gesture and speech (a multimodal approach) felt natural to the participants, compared with using a single input modality by itself.

2.3.2 Multimodal Interactions. In contrast to unimodal interaction, multimodal interaction refers to the combination of two or more unimodal interactions. Recent articles support the use of multimodal interaction to improve the user's workflow, communication, and performance [21]. In prior work, Rodriguez et al. [49], based on participant feedback, suggested using gestures and speech for unimodal and multimodal interactions. Some studies have shown that using multimodal interaction in communication apps, as in Meet me in VR [1] enhanced the experience and made participants feel more engaged while meeting virtually. Bittner et al. [8] also

demonstrated that using multimodal interactions when one person is creating a photobook allows the individual to progress fluidly.

3 Motivation and Research Questions

Flow has been characterized as the immersion or concentration that a performer has while involved in a specific experience [6]. A person's flow during an activity is important because it is an indication of the person's concentration on that activity, that the person has clear goals, and that they are enjoying the activity [6, 31]. Since an activity can be done with just about anything, this flow can therefore be used in several disciplines, from games [44, 55, 56] and exercise [16] to more serious disciplines, such as surgery [58]. To quantify these experiences, Jackson et al. [30] developed the Flow State Scale (FSS), a validated psychometric tool that resembles a Likert questionnaire, and considers the nine dimensions of flow: Challenge-Skill Balance, Action-Awareness Merging, Clear Goals, Unambiguous Feedback, Concentration on the Task, Sense of Control, Loss of Self-Consciousness, Transformation of Time, and Autotelic Experience.

As creative drawing is a process, we can consider the creative flow ([19], [20]) to measure artists' focus while sketching. In VR, artists can experience creative drawing at a new level. When a drawing application in VR is combined with multimodal interactions, then we need to consider how the flow is affected by these combinations.

A key question is how the flow is affected when users are immersed in a virtual environment and actively engaged in drawing. A prior study by Rodriguez et al. [49] examined VR sketching in OpenBrush using VR controllers as the only input method. Feedback from that experiment indicated that participants felt there were too many choices reducing performance, as users had to shift their attention away from the creative process to locate the appropriate brush or command. This motivated further investigation into how alternative input modalities (such as controllers, microgestures, or augmenting these inputs with speech) may reduce such interruptions and support creative flow. Therefore, we pose the following research questions:

- RQ_1. Does the base input, controller vs microgesture, affect user performance and flow during VR sketching?
- RQ_2. Does the use of Unimodal versus Multimodal interactions affect flow during VR sketching?
- RQ_3. Which dimensions of creative flow are most affected by the interaction technique used during VR sketching?

4 Evaluated Interfaces

The study was conducted using a custom VR sketching application, NUI-Sketch [1], developed for this experiment. The system allowed participants to draw free-form 3D strokes in VR and supported common sketching operations such as drawing, erasing, undoing, redoing, selecting, deleting, scaling, and changing color. The application was designed to provide a consistent sketching environment across all four input conditions, while varying only the available input modalities.

In the controller-based conditions, participants interacted with the sketching system using handheld VR controllers. Controller input supported the core sketching actions required for the task, including drawing, undo, redo, selection and manipulation, color selection, erasing, deleting selected strokes, moving selected strokes, and scaling selected strokes. Figure 2 (left) shows the action maps corresponding to the controller buttons and triggers for drawing and manipulation modes.

[1] apk available upon request

Table 1: Speech commands available.

Item	Command	Description
1	undo	does away with the last command
2	redo	restores last command after undo
3	increase	increases the stroke size
4	decrease	decreases the stroke size
5	erase	switches to the eraser
6	brush	switches to the brush
7	menu	brings up the color menu

In the microgesture-based conditions, microgesture input was intended to provide a more direct and embodied interaction style, allowing participants to sketch and manipulate content using hand movements (see Figure 2). The right shows the six microgestures used in our study (pinch, tab, and four directional swipes), and the middle figure presents the microgesture action map, illustrating the corresponding microgesture commands for drawing and manipulation modes.

In the multimodal conditions, speech was added as a complementary input modality. Speech was not used as a standalone condition and was paired with controller or microgesture input to support command-based actions that could reduce menu navigation and interaction overhead. Originally, we considered 20 speech commands. However, a pilot study revealed that 20 speech commands were too many for participants to remember. Therefore, we reduced the speech command set to seven commands, most of which users regularly use on desktop systems. These commands, along with their descriptions, are presented in Table 1.

4.1 Apparatus

An ASUS Notebook G16 with 32GB of onboard RAM, running Windows 11 Pro 64-bit version 10.0.26200 was used to run the VR application. The notebook also housed an NVIDIA GeForce RTX 4070 Laptop GPU. A custom sketching application, called NUI-Sketch, was developed in Unity 3D version 6000.0.58f2 utilizing the Meta XR All-in-One SDK version 77.0.0 and OpenXR Plugin version 1.15.1 with the C# scripts edited in Visual Studio. The app ran in Unity 3D version 6000.0.58f2, along with the Meta Horizon Link, version 80.0.0.306.552, which was used to connect to the VR Headset. A Meta Quest 3 and a Meta Quest Link cable were used to connect to the notebook. OBS Studio was used to record what the participants were viewing on the HMD. For this study we used a Wizard of Oz paradigm to simulate the system's speech responses, allowing us to examine speech as an interaction modality under controlled recognition conditions, without introducing recognition failures or recognizer-dependent variability that could differ across participants [51]. The VR environment where the participants moved around freely to sketch was approximately 3 m by 2.5 m.

5 Methodology

5.1 Participants

A total of 24 participants were recruited from the local university. Initially, the total number of participants was 30, but 6 were removed due to equipment malfunction. The participants' ages ranged from 19 to 44 ($M = 24.29, SD = 5.84$). Twelve participants identified as female, eleven as male, and one as non-binary. All participants self-identified as right-handed. For this study, no participant had high susceptibility to cybersickness. The study protocol was approved by the Institutional Review Board (IRB) under Protocol #4007.

5.2 Experimental Design

The study employed a $2*2$ within-subject factorial design with *base input* (Controller, Microgesture) and *speech augmentation* (Unimodal, Multimodal) as independent variables (see Figure 1). These factors yield four experimental conditions: *Controller* (C), *Controller+Speech* (C+S), *Microgesture* (MG), and *Microgesture+Speech* (MG+S). Each participant was asked to complete a sketching task in VR under all four conditions, and the order of the conditions was counterbalanced across participants.

5.3 Tasks

Our experiment included four tasks that participants had to complete. We followed the same methodology introduced by Rodriguez et al. [49]. In each task, the participant was asked to sketch a dog in 3D. We intentionally used the same target category across all four conditions to maintain a consistent task goal and support comparison of the interaction techniques. Participants were not asked to reproduce a fixed reference; instead, they were free to create their own envisioned 3D dog as they saw fit. They could spend up to ten (10) minutes on each sketch, but if they were done before the time limit, they could inform the researcher that they were done. During the task, the participant was free to erase, redraw, and create new strokes to better represent the image they had in mind.

5.4 Procedure

Before the experiment, all participants were shown how to walk around the VR playground, detect when they were near the wall, and use the controllers and microgestures on the Meta Quest 3. This was done for all participants, regardless of whether they had any previous experience using VR, so that no participant could have an advantage over another who had not used a VR headset before. They practiced in a mini-session directly on the application using the Meta Quest 3 controllers, microgestures, and speech, for up to 10 minutes. The participants were asked to sketch anything to gain familiarity with the system. The researcher then asked them to use different functions, such as erase, undo, or redo, to help them learn the basic usage of the sketching app. The participants would practice with the *Controller*, followed by *Controller+Speech*, then *Microgesture*, and finally *Microgesture+Speech*. At the beginning of each condition, participants were informed which input methods were available for that condition: controller or microgesture input, with or without speech. Because each participant would spend up to 40 minutes completing the tasks, all participants agreed per the study inclusion criteria, that they did not suffer from cybersickness. Participants were also able to stop the experiment at any time.

Once the practice was completed, the study began using the four previously mentioned conditions. At the beginning of each task, the researcher would remind the participant of the total time they had to sketch the dog. If the 10-minute mark was reached, the researcher would inform the participant that the time was up; if the participant finished before the 10-minute limit, the participant would alert the researcher. After each condition, the researcher would ask the participant to complete three questionnaires: a Flow State Scale questionnaire, a VR Presence questionnaire, and a System Usability Scale questionnaire. Once the last condition was completed, a post-study interview was conducted between the researcher and the participant (questionnaires are provided in the supplemental materials).

5.5 Evaluation Metrics

To evaluate both objective and subjective user performance and experience, we collected behavioral, usability, presence, and flow-related measures.

Action Counts (AC). AC is a metric for sketch evaluation [5] and refers to the number of times a participant opened or activated the Menu/Color function during a trial. We operationalized this measure by counting detected openings of the color-panel interface in the recorded first-person VR videos. We focused on color-panel activations because this interface element had a clear and consistent visual signature across all four conditions.

Task Completion Time (TCT). The interval between the start of the task and successful task completion. To prevent excessively long sessions from disproportionately influencing the results, task completion time was capped at 10 minutes.

Flow Experience (FSS-2). Participants completed the 36-item Flow State Scale-2 (FSS-2) [30, 57] (see the FSS-2 questionnaire, provided in the supplemental materials), which measures the extent to which users experience a state of flow during an activity. The scale assesses nine dimensions of flow, including challenge–skill balance, action–awareness merging, unambiguous feedback, sense of control, clear goals, concentration on the task, loss of self-consciousness, transformation of time, and autotelic experience. Responses were recorded on a 5-point Likert scale, with higher scores indicating a stronger flow experience.

System Usability Scale (SUS). Participants completed the 10-item System Usability Scale (SUS) [3, 12, 35] (see the System Usability Scale (SUS) questionnaire, provided in the supplemental materials), a widely used subjective measure of overall system usability. Items are rated on a 5-point Likert scale and combined into a single score ranging from 0 to 100, where higher scores indicate greater perceived usability.

Presence (PQ). Participants completed a 24-item subset of the Presence Questionnaire (PQ) [52, 61] (see the Presence (PQ) questionnaire, provided in the supplemental materials), which measures

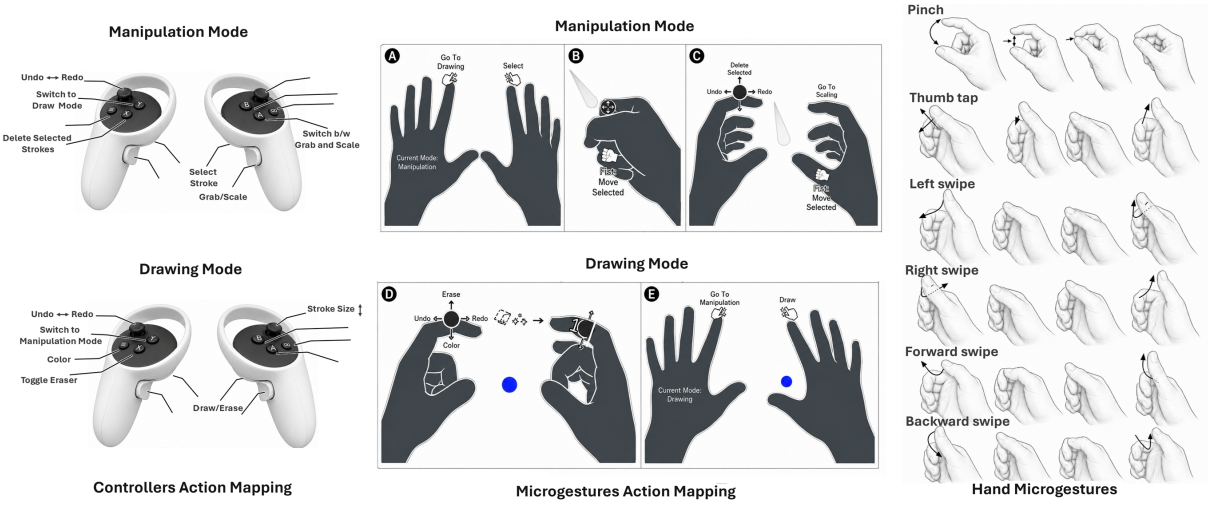


Figure 2: Interaction mappings. (left) The controller action map shows the button and trigger mappings for drawing and manipulation modes. (middle) the microgesture action map shows the corresponding hand-based commands for drawing, manipulation, selection, movement, and scaling on main and secondary hands. (right) The hand microgesture panel illustrates the thumb-tap and thumb-swipe microgestures used for directional input.

the extent to which users feel present or experience a sense of "being there" within the virtual environment. Higher scores indicate a stronger sense of presence in the virtual environment.

Preference and Qualitative Feedback. After completing all conditions, participants answered seven semi-structured interview questions about their experiences with the interaction techniques. The questions compared single-input conditions, using either the controller or microgestures, with multimodal conditions, in which speech was combined with the controller or microgestures. Participants were asked about ease of use, perceived speed, comfort over time, creativity and flexibility, future preference for single versus combined input, and which main input method better supported task flow and faster completion. The interview also asked whether participants' workflow felt natural within each task and as they progressed across tasks.

6 Results

This analysis tested the main effect of base input, the main effect of speech augmentation, and the base input × speech augmentation interaction, which examined whether the effect of adding speech differed between controller and microgesture input.

For each dependent measure, we conducted a two-way repeated-measures ANOVA with base input and speech augmentation as within-subject factors. We also ran the non-parametric Aligned Rank Transform (ART) procedure [62], which supports factorial testing of main effects and interactions on ranked data. Residual checks showed that Flow and Presence met the normality assumption, whereas SUS and task completion time did not. Accordingly, we use the parametric ANOVA results for Flow and Presence, and treat the ART results as the primary inferential tests for SUS and task time, with the corresponding ANOVA results reported only for comparison.

6.1 Effects of Base Input

Table 2 presents the statistical analysis for the effect of base input. Figure 3 shows the comparison of base input across our dependent measures.

Activation counts. Active action counts were similar between the *Controller* and *Microgesture* conditions, with slightly higher counts for *Controller* ($M = 6.77$) than for *Microgesture* ($M = 6.34$). As shown in Table 2, this difference was not significant, $F(1, 21) = 0.17$, $p = .684$; ART $p = .626$. Thus, the base input method alone did not reliably affect how often participants opened or activated the color/menu function.

Task Time. Participants completed the task fastest with C+S ($Md = 406\ s$), followed by C ($Md = 446\ s$), MG ($Md = 494\ s$), and MG+S ($Md = 556\ s$). The 10-minute cap was reached by 16.7%, 25.0%, 16.7%, and 25.0% participants, respectively. When comparing the two base inputs, *Controller* led to shorter task times than the *Microgesture* condition (425 s vs. 489 s), averaged across both speech conditions. This difference was significant, $F(1, 23) = 6.98$, $p = 0.015$, $\eta^2_G = 0.052$; the ART test supported the same conclusion, $F = 6.10$, $p = 0.021$. Task time was the only dependent measure for which the two base inputs differed significantly.

Flow. For input modality, the *Controller* and *Microgesture* conditions produced very similar Flow scores. Averaged across speech augmentation, the marginal means were 4.08 for Controller and 4.15 for *Microgesture* on the 1–5 FSS scale. Within the *Controller* conditions, mean Flow was $M = 4.01$ ($SD = 0.48$) for C and $M = 4.16$ ($SD = 0.42$) for C+S. Within the *Microgesture* conditions, mean Flow was $M = 4.03$ ($SD = 0.38$) for MG and $M = 4.28$ ($SD = 0.37$) for MG+S. The main effect of input was not significant, $F(1, 23) = 0.68$, $p = 0.419$, $\eta^2_G = 0.007$ (ART $p = 0.585$).

Table 2: Main effect of *input* (*Controller* vs. *Microgesture*) per measure, averaged over speech. RM-ANOVA and Aligned Rank Transform (ART); df = 1, 23. Bold p-values are significant ($p < 0.05$). Task-time effects are exploratory (time right-censored at the 600 (seconds) cap).

	RM-ANOVA			ART	
Measure	F	p	η^2_G	F	p
Action counts	0.17	0.684	0.003	0.24	0.626
Task time (s)	6.98	**0.015**	0.052	6.10	**0.021**
Flow (FSS-2)	0.68	0.419	0.007	0.31	0.585
Usability (SUS)	0.01	0.932	0.000	0.06	0.808
Presence (PQ)	0.33	0.573	0.002	0.37	0.551

Usability. SUS scores were almost identical between *Controller* and *Microgesture*. Averaged across speech augmentation, the marginal means were 79.11 for controller and 78.96 for microgesture input on the 0–100 SUS scale. Within the *Controller* conditions, SUS scores were $M = 78.65$ ($SD = 12.14$) for C and $M = 79.58$ ($SD = 12.55$) for C+S. Within the *Microgesture* conditions, SUS scores were $M = 76.15$ ($SD = 11.40$) for MG and $M = 81.77$ ($SD = 10.15$) for MG+S. The main effect of input was not significant, $F(1, 23) = 0.01$, $p = 0.932$, $\eta^2_G = 0.000$ (ART $p = 0.808$).

Presence. PQ scores were very similar between *Controller* and *Microgesture*. Averaged across speech augmentation, the marginal means were 5.11 for Controller and 5.17 for *Microgesture* on the 1–7 PQ scale. Within the *Controller* conditions, Presence scores were $M = 5.07$ ($SD = 0.72$) for C and $M = 5.16$ ($SD = 0.58$) for C+S. Within the *Microgesture* conditions, Presence scores were $M = 5.04$ ($SD = 0.52$) for MG and $M = 5.29$ ($SD = 0.56$) for MG+S. The main effect of input was not significant, $F(1, 23) = 0.33$, $p = 0.573$, $\eta^2_G = 0.002$ (ART $p = 0.551$).

6.2 Effects of Speech Augmentation

Table 3 presents the statistical analysis for the effect of speech augmentation. Figure 4 shows the comparison of Unimodal and Multimodal interactions across our dependent measures.

Activation counts. Speech augmentation was associated with a higher number of detected activations, increasing from $M = 5.00$ in the Unimodal conditions to $M = 8.11$ in the multimodal conditions. As shown in Table 3, this main effect of speech was significant, $F(1, 21) = 25.17$, $p < .001$. This indicates that adding speech led participants to activate the menu/color function more frequently.

Task Time. Speech augmentation did not significantly affect task completion time, $F(1, 23) = 1.87$, $p = 0.184$, $\eta^2_G = 0.007$. The ART analysis led to the same conclusion, $F = 1.85$, $p = 0.186$. Thus, adding speech did not provide evidence of faster task completion.

Flow. Mean flow score rose from $M = 4.02$ (Unimodal) to $M = 4.22$ (Multimodal), a 0.20-point gain on the 1–5 scale. The main effect of speech augmentation was significant, $F(1, 23) = 30.28$, $p < .001$, $\eta^2_G = 0.059$ (ART $F = 27.23$, $p < 0.001$). Although statistically significant, the magnitude of the observed improvement was modest.

Table 3: Main effect of *speech augmentation* (Unimodal vs. Multimodal) per measure, averaged over input. RM-ANOVA and Aligned Rank Transform (ART); df = 1, 23. Bold p-values are significant ($p < 0.05$).

	RM-ANOVA			ART	
Measure	F	p	η^2_G	F	p
Action counts	25.17	**<0.001**	0.126	24.86	**<0.001**
Task time (s)	1.87	0.184	0.007	1.85	.186
Flow (FSS-2)	30.28	**<0.001**	.059	27.23	**<0.001**
Usability (SUS)	4.61	**0.043**	.020	6.25	**0.020**
Presence (PQ)	14.62	**<0.001**	.020	17.66	**<0.001**

Usability. SUS scores were significantly higher in the Multimodal conditions than in the Unimodal conditions. Mean SUS increased from $M = 77.40$ without speech augmentation to $M = 80.68$ with speech augmentation, corresponding to a 3.28-point gain on the 0–100 SUS scale. This improvement moves the system from the upper end of the good" range toward the excellent" range. The main effect of speech augmentation was significant, $F(1, 23) = 4.61$, $p = .043$, $\eta^2_G = .020$; the ART test also confirmed this effect, $F = 6.25$, $p = 0.020$. The magnitude of the observed improvement was modest. Because the SUS residuals were non-normal (Shapiro–Wilk $p < 0.001$), the ART result is treated as the primary test.

Presence. PQ scores were significantly higher in the Multimodal conditions than in the Unimodal conditions. Mean Presence increased from $M = 5.05$ without speech augmentation to $M = 5.22$ with speech augmentation, reflecting a modest gain on the 1–7 PQ scale. This main effect of speech augmentation was significant, $F(1, 23) = 14.62$, $p < 0.001$, $\eta^2_G = 0.020$; the ART test showed the same pattern, $F = 17.66$, $p < 0.001$.

6.3 Base Input × Speech Augmentation

The input × speech interaction was non-significant across all dependent measures, indicating that the effect of speech augmentation did not significantly differ between *Controller* and *Microgesture* conditions. Table 4 shows the statistical analysis. For flow, the interaction was not significant, $F(1, 23) = 1.23$, $p = 0.280$, $\eta^2_G = 0.003$; the ART analysis led to the same conclusion, $F = 1.27$, $p = 0.272$. However, the cell means suggest a consistent trend. As shown in Figure 5, the increase from Unimodal to Multimodal interaction was descriptively larger for microgestures than for the controllers. The *Microgesture + Speech* condition also produced the highest mean scores for flow (4.28), usability (81.77), and presence (5.29). However, because the interaction was not statistically significant, these descriptive differences do not provide evidence that microgesture input benefited more from speech than the controller input did.

6.4 Nine Dimensions of Flow

To better understand the source of observed flow differences, we analyzed the nine flow sub-dimensions separately (Table 5). Following Benjamini-Hochberg correction for multiple comparisons ($q = .05$), only three dimensions remained significant: Action–Awareness Merging ($p_{adj} = 0.021$), Unambiguous Feedback ($p_{adj} = 0.021$), and

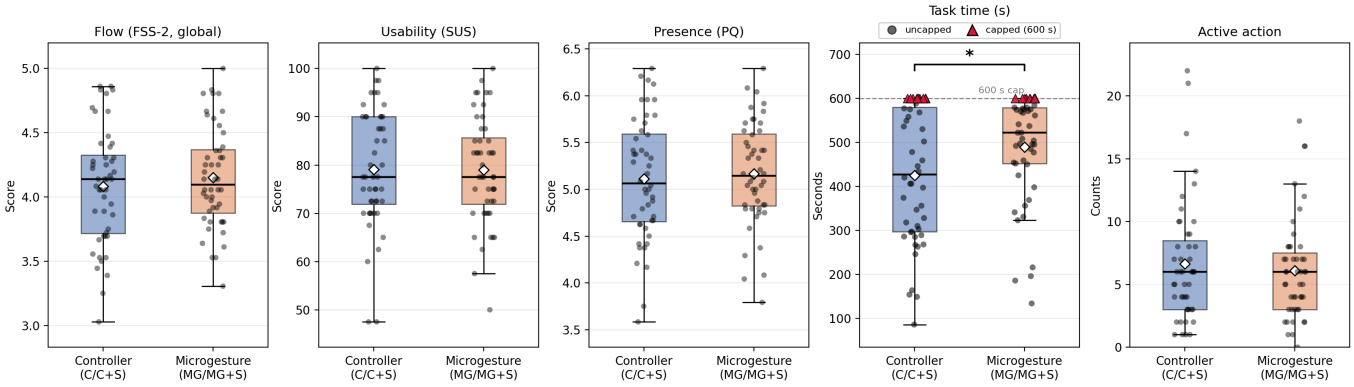


Figure 3: Comparison of *base input* (*Controller* vs *Microgesture*) across FSS, SUS, PQ, TCT, and actions. TCT was capped at 600 s; red triangles indicate participants who reached the time cap.

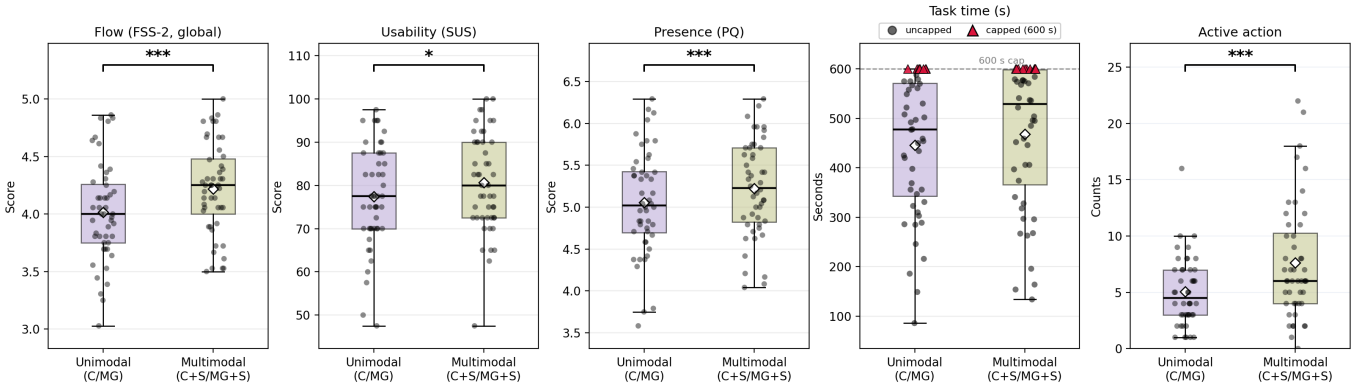


Figure 4: Comparison of *speech augmentation* (Unimodal vs Multimodal) across FSS, SUS, PQ, TCT, and AC.

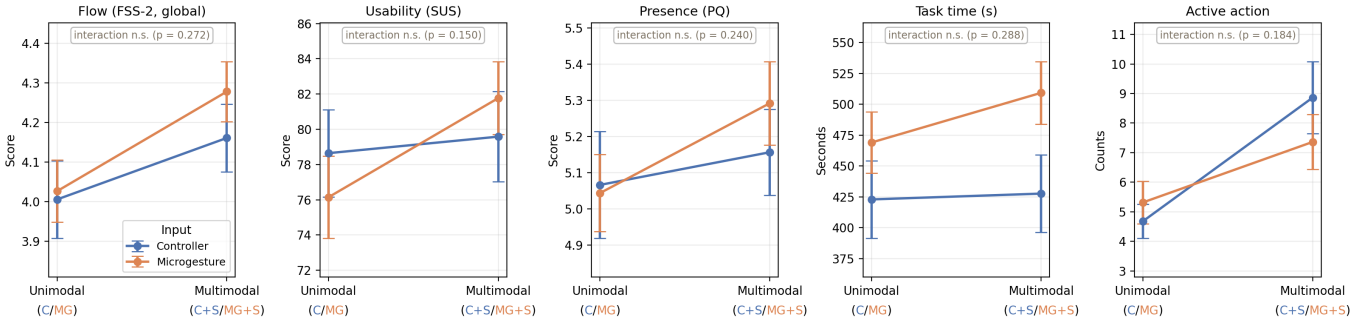


Figure 5: Input × speech interaction for FSS, SUS, PQ, TCT, and AC. Each panel plots the controller (blue) and microgesture (orange) input means from Unimodal to Multimodal.

Sense of Control ($p_{\text{adj}} = 0.021$). The remaining six dimensions did not differ significantly across conditions.

6.5 Interviews

We used thematic analysis [17] to analyze participants' feedback collected through seven semi-structured interview questions. Across 24 coded interviews, participants' answers were split into codable segments, resulting in 337 segments. Each segment was assigned descriptive codes, which resulted in the emergence of six themes: (1) *preference was shaped by habit and familiarity.* 70% of participants described preferences that were influenced by prior experience with controllers, pens, or single-input interaction, rather than by speech failing as an input method. (2) *Speech helps by removing a navigation step.* 65% of participants described speech as useful because it removed intermediate interaction steps, such as opening menus, searching for buttons, changing colors, undoing,

Table 4: *Input* × *speech augmentation* interaction per measure. RM-ANOVA and Aligned Rank Transform (ART); df = 1, 23.

	RM-ANOVA			ART	
Measure	F	p	η^2_G	F	p
Action counts	1.92	0.181	0.17	1.89	0.184
Task time (s)	0.79	0.383	0.004	1.18	0.288
Flow (FSS-2)	1.23	0.280	0.003	1.27	0.272
Usability (SUS)	2.33	0.141	0.011	2.22	0.150
Presence (PQ)	1.28	0.270	0.004	1.45	0.240

Table 5: Friedman test results for the nine FSS-2 subdimensions with Benjamini–Hochberg FDR correction (q = 0.05). Dimensions that remained significant after correction are shown in bold.

Dimension	Raw p	p_{adj}
Action–Awareness	0.004	**0.021**
Unambiguous Feedback	0.006	**0.021**
Sense of Control	0.007	**0.021**
Clear Goals	0.026	0.058
Challenge–Skill Balance	0.535	0.564
Concentration	0.564	0.564
Loss of Self-Consciousness	0.333	0.428
Transformation of Time	0.294	0.428
Autotelic Experience	0.234	0.421

or erasing. (3) *base inputs trade off precision against immersion.* 70% of participants discussed a tradeoff between the base input modalities: controllers were perceived as more precise and reliable, whereas microgestures felt more natural, immersive, or playful but were sometimes less accurate or more prone to accidental input. (4) *multimodality helps coordination but does not add creativity.* 55% of participants indicated that multimodal interaction improved coordination and flexibility. However, they generally described it as streamlining the workflow rather than directly increasing creativity. (5) *The benefit of speech grows with practice.* 45% of participants reported that the benefits of speech and multimodal input increased with practice, suggesting that initial unfamiliarity may have limited their use of speech commands. (6) *combining more than two input modalities was not fevered.* 35% of participants suggested that combining one manual modality with speech was useful, but adding too many modalities could become confusing.

7 Discussion

Does the base input, controller vs microgesture, affect user performance and flow during VR sketching? *(RQ_1).* Our results show that the base input modality affected performance, but not overall creative flow, usability, or presence. Controller-based input led to significantly shorter task completion times than microgesture-based input, suggesting that controllers remain more efficient for completing the sketching task [15]. This result is consistent with participants' qualitative feedback and findings from a prior study comparing controller and hand gestures [47], in which controllers were often described as more precise and reliable, while microgestures were perceived as more natural, immersive, and playful. However, the lack of significant differences in flow indicates that faster performance did not necessarily translate into a stronger creative experience.

Does the use of unimodal versus multimodal interactions affect flow during VR sketching? (RQ_2) The findings indicate that multimodal interaction, specifically adding speech to the base input modality, improved the subjective sketching experience [46]. Speech augmentation significantly increased flow, usability, and presence compared with unimodal interaction, although the magnitude of these improvements was modest. This suggests that speech helped participants remain engaged in the creative process by reducing the need to interrupt drawing for command-based actions, such as undoing, erasing, changing brush size, or opening menus [11, 33]. Interview results support this interpretation: many participants described speech as useful because it removed intermediate interaction steps such as menu navigation and button searching. However, speech did not significantly reduce task completion time, indicating that multimodality improved the quality of the experience more than objective speed. Therefore, multimodal interaction is useful for supporting creative flow.

Flow dimensions most affected by interaction technique. *(RQ_3).* The analysis of the nine FSS-2 [30] dimensions showed that the interaction technique did not affect all aspects of flow equally. Only three dimensions remained significant: Action-Awareness Merging, Unambiguous Feedback, and Sense of Control. This suggests that the main benefit of multimodal interaction was not a general increase across all components of flow, but rather a more specific improvement in how smoothly users could act, understand system responses, and feel in control while sketching. These dimensions are directly related to interaction design: when users can issue commands through speech without leaving the drawing posture, they may experience fewer interruptions between intention and action. At the same time, clear command execution and reliable feedback can help users maintain control.

8 Limitations

Our study used a Wizard-of-Oz implementation for speech input to maintain consistent recognition behavior across participants and conditions and to isolate the effect of adding speech as an interaction modality. As a result, the findings represent speech augmentation under controlled recognition conditions and do not capture recognition errors or recognizer-dependent latency that may occur in deployed speech systems. Such factors could affect interaction flow, perceived usability, and the magnitude of the benefits observed for speech-enabled conditions. A follow-up study using an implemented speech recognizer would provide an important opportunity to examine how recognition accuracy and response latency influence the interaction experience.

Another limitation was the participants' age. The study was open to the general public, and participants had to reside locally and be between 18 and 65 years old. However, it was conducted at

a local university, where one limitation was that most participants were students, with ages ranging from 19 to 44, which does not give us insight into whether participants over the age of 44 would have different opinions.

Additionally, the interaction techniques were evaluated within a single application and hardware setup. This study did not examine other input techniques, such as stylus-based interaction, collaborative workflows, or more open-ended creative tasks. Future work will expand this investigation by exploring a wider range of sketching activities and recruiting participants outside academia. We also plan to examine additional input modalities to understand better how VR sketching systems can support different skill levels, personal preferences, and evolving creative practices over time.

Finally, using the same sketching prompt across conditions may constrain the scope of the findings. Although participants created their own representations rather than copying a fixed reference, repeatedly sketching the same semantic object is less open-ended than many artistic workflows and may introduce familiarity, learning, or boredom over the course of the session.

9 Conclusion

In this study, 24 participants completed VR sketching tasks using controller and microgesture input, with and without speech augmentation. The results showed that the choice of input modality influences the creative process. Multimodal interaction also affects the users' creative flow, sense of presence, and overall experience. Among the interaction techniques evaluated, the combination of microgestures and speech (MG+S) produced the highest descriptive scores for flow, system usability, and sense of presence. However, the interaction between the base input modality and speech was not significant, providing no evidence that speech benefited microgesture input more than controller input did. In addition, controller-based interaction combined with speech enabled participants to complete tasks more quickly. These findings suggest that designing VR sketching systems around speed alone may overlook other aspects of the experience that are important for creative work. Accordingly, speech augmentation may offer strong potential to support a deeper state of flow and a sense of engagement when creating art.

Acknowledgments

ONR N00014-24-1-2214. National Science Foundation NSF 2444132, NSF 2439474, NSF 2327569, NSF 2238313, and NSF 2223432. We also thank Sophia XR for partial travel support (sophiaxr.com), an Agentic AI and Spatial Extended Reality company.

References

[1] Katarzyna Abramczuk, Zbigniew Bohdanowicz, Bartosz Muczyński, Kinga H. Skorupska, and Daniel Cnotkowski. 2023. Meet me in VR! Can VR space help remote teams connect: A seven-week study with Horizon Workrooms. *International Journal of Human-Computer Studies* 179 (2023), 103104. doi:10.1016/j.ijhcs.2023.103104

[2] Rahul Arora, Mayra Donaji Barrera Machuca, Philipp Wacker, Daniel Keefe, and Johann Habakuk Israel. 2023. Introduction to 3D sketching. In *Interactive Sketch-Based Interfaces and Modelling for Design*. River Publishers, New York, USA, 151–177. doi:10.1201/9781003360650

[3] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Humanâ€“Computer Interaction* 24, 6 (2008), 574–594. doi:10.1080/10447310802205776

[4] Baroque Software. 2025. *VR Sketch*. https://vrsketch.eu/

[5] Mayra Donaji Barrera Machuca, Johann Habakuk Israel, Daniel F Keefe, and Wolfgang Stuerzlinger. 2023. Toward more comprehensive evaluations of 3D immersive sketching, drawing, and painting. *IEEE Transactions on Visualization and Computer Graphics* 30, 8 (2023), 4648–4664. doi:10.1109/TVCG.2023.327629

[6] Lawrence A. Beck. 1992. Csikszentmihalyi, Mihaly. (1990). Flow: The Psychology of Optimal Experience. *Journal of Leisure Research* 24, 1 (1992), 93–94. doi:10.1080/00222216.1992.11969876

[7] Yulong Bian, Chenglei Yang, Fengqiang Gao, Huiyu Li, Shisheng Zhou, Hanchao Li, Xiaowen Sun, and Xiangxu Meng. 2016. A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Personal and Ubiquitous Computing* 20, 5 (2016), 821–832. doi:10.1007/s00779-016-0953-5

[8] Sara-Jane Bittner, Robert Andreas Leist, László Kopácsi, Omair Shahzad Bhatti, Abdulrahman Mohamed Selim, Michael Barz, and Daniel Sonntag. 2025. Interactive Multimodal Photobook Co-Creation in Virtual Reality. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 146–151. doi:10.1145/3708557.3716355

[9] Richard A. Bolt. 1980. "Put-That-There": Voice and Gesture at the Graphics Interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques* (Seattle, Washington, USA) *(SIGGRAPH '80)*. Association for Computing Machinery, New York, NY, USA, 262–270. doi:10.1145/800250.807503

[10] Alexandra Bonnici, Kenneth P Camilleri, Alexandra Bonnici, and Kenneth P. Camilleri. 2023. *Interactive Sketch-Based Interfaces and Modelling for Design* (1 ed.). River Publishers, Milton. doi:10.1201/9781003360650

[11] Stephen Brade, Sam Anderson, Rithesh Kumar, Zeyu Jin, Anh Truong, Naomi Yamashita, Vanessa Evers, Koji Yatani, Xianghua (Sharon) Ding, Bongshin Lee, Marshini Chetty, and Phoebe Toups-Dugas. 2025. SpeakEasy: Enhancing Text-to-Speech Interactions for Expressive Content Creation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–19. doi:10.1145/3706598.3714263

[12] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7. doi:10.1201/9781498710411-35

[13] Adrien Chaffangeon Caillet, Alix Goguey, and Laurence Nigay. 2023. 3d selection in mixed reality: Designing a two-phase technique to reduce fatigue. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, IEEE ISMAR, Sydney, Australia, 800–809. doi:10.1109/ISMAR59233.2023.00095

[14] Tatiana Chemi. 2016. The experience of flow in artistic creation. In *Flow experience: Empirical research and applications*. Springer, Switzerland, 37–50. doi:10.1007/978-3-319-28634-1_3

[15] Chen Chen, Matin Yarmand, Zhuoqun Xu, Varun Singh, Yang Zhang, and Nadir Weibel. 2022. Investigating Input Modality and Task Geometry on Precision-first 3D Drawing in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Piscataway, NJ, USA, 384–393. doi:10.1109/ISMAR55827.2022.00054

[16] Anand Chockalingam and Kavin Anand. 2021. Encouraging Flow : A State Of Calm and Deep Focus May Improve Exercise Compliance in Cardiac Rehabilitation. *Missouri medicine* 118, 6 (2021), 556–560. https://pmc.ncbi.nlm.nih.gov/articles/PMC8672947/

[17] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298. doi:10.1080/17439760.2016.1262613

[18] Genevieve M. Cseh, Louise H. Phillips, and David G. Pearson. 2015. Flow, affect and visual creativity. *Cognition and Emotion* 29, 2 (2015), 281–291. doi:10.1080/02699931.2014.913553

[19] Mihaly. Csikszentmihalyi. 1996. *Creativity : flow and the psychology of discovery and invention* (1st ed. ed.). HarperCollinsPublishers, New York.

[20] Charlotte L. Doyle. 2017. Creative Flow as a Unique Cognitive Process. *Frontiers in psychology* 8 (2017), 1348–. doi:10.3389/fpsyg.2017.01348

[21] Ryan Khushan Ghamandi, Ravi Kiran Kattoju, Yahya Hmaiti, Mykola Maslych, Eugene Matthew Taranta, Ryan P. McMahan, and Joseph LaViola. 2024. Unlocking Understanding: An Investigation of Multimodal Communication in Virtual Reality Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 602, 16 pages. doi:10.1145/3613904.3642491

[22] Google. 2016. *Tilt Brush*.

[23] Gravity Sketch. 2017. *Gravity Sketch*. https://gravitysketch.com/

[24] Mark D. Gross and Ellen Yi-Luen Do. 1996. Ambiguous intentions: a paper-like interface for creative design. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*. ACM, New York, NY, USA, 183–192. doi:10.1145/237091.237119

[25] Sang-Lin Han, Myounga An, Jerry J Han, and Jiyoung Lee. 2020. Telepresence, time distortion, and consumer traits of virtual reality shopping. *Journal of Business Research* 118 (2020), 311–320. doi:10.1016/j.jbusres.2020.06.056

[26] Jeffrey T Hansberger, Chao Peng, Shannon L Mathis, Vaidyanath Areyur Shanthakumar, Sarah C Meacham, Lizhou Cao, and Victoria R Blakely. 2017. Dispelling the gorilla arm syndrome: the viability of prolonged gesture interactions. In *International conference on virtual, augmented and mixed reality*. Springer, Springer, Vancouver, Canada, 505–520. doi:10.1007/978-3-319-57987-0_41

[27] Samory Houzangbe, Sylvain Fleury, Dimitri Masson, David Antonio Gómez Jáuregui, Jérémy Legardeur, Nadine Couture, and Simon Richir. 2023. Exploration of Physiological Arousal in Divergent and Convergent Thinking using 2D screen and VR Sketching Tools. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, IEEE, Cambridge, MA, USA, 1–8. doi:10.1109/ACII59096.2023.10388141

[28] Meng-Hsuan Huang and Saiau-Yue Tsau. 2018. A Flow Experience Analysis on the Virtual Reality Artwork: La Camera Insabbiata. In *Proceedings of the International Conference on Machine Vision and Applications* (Singapore, Singapore) *(ICMVA '18)*. Association for Computing Machinery, New York, NY, USA, 51–55. doi:10.1145/3220511.3220514

[29] Icosa Foundation. 2021. *Open Brush*. https://openbrush.app/

[30] Susan A. Jackson and Herbert W. Marsh. 1996. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of sport & exercise psychology* 18, 1 (1996), 17–35. doi:10.1123/jsep.18.1.17

[31] Misha Jan. 2024. *Flow State for Mental Health: Empathy and Creativity*. Psychology Today. Blog post in "Love Is An Action". https://www.psychologytoday.com/us/blog/love-is-an-action/202406/flow-state-for-mental-health-empathy-and-creativity

[32] Daniel F. Keefe, Daniel Acevedo Feliz, Tomer Moscovich, David H. Laidlaw, and Joseph J. LaViola. 2001. CavePainting: a fully immersive 3D artistic medium and interactive experience. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics (I3D '01)*. Association for Computing Machinery, New York, NY, USA, 85–93. doi:10.1145/364338.364370

[33] Yea-Seul Kim, Mira Dontcheva, Eytan Adar, and Jessica Hullman. 2019. Vocal Shortcuts for Creative Experts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. doi:10.1145/3290605.3300562

[34] Ladd Labb Interactive, Inc. 2024. *Gesture VR*. https://www.gesture-vr.com/

[35] James R. Lewis. 2018. The System Usability Scale: Past, Present, and Future. *International Journal of Human – Computer Interaction* 34, 7 (2018), 577–590. doi:10.1080/10447318.2018.1455307

[36] Guangchuan Li, David Rempel, Yue Liu, and Carisa Harris-Adamson. 2020. The design and assignment of microgestures to commands for virtual and augmented reality tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 64. SAGE Publications Sage CA: Los Angeles, CA, SAGE Publications, Los Angeles, CA, 2061–2063. doi:10.1177/10711813206414

[37] Guangchuan Li, David Rempel, Yue Liu, Weitao Song, and Carisa Harris Adamson. 2021. Design of 3d microgestures for commands in virtual reality or augmented reality. *Applied Sciences* 11, 14 (2021), 6375. doi:10.3390/app11146375

[38] Xiang Li, Wei He, and Per Ola Kristensson. 2025. Evaluating the usability of microgestures for text editing tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 32 (2025), 2020–2033. Issue 2. doi:10.1109/TVCG.2025.3642050

[39] Ling Luo, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song, and Yulia Gryaditskaya. 2023. 3D VR Sketch Guided 3D Shape Prototyping and Exploration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Paris, France, 9233–9242. doi:10.1109/ICCV51070.2023.00850

[40] I. Scott MacKenzie. 2024. *Human-computer interaction : an empirical research perspective* (second edition. ed.). Morgan Kaufmann, Cambridge, MA. doi:10.1016/C2022-0-02755-0

[41] Hongyu Mao, Mar Gonzalez-Franco, Vrushank Phadnis, Eric J Gonzalez, and Ishan Chatterjee. 2025. Restfulraycast: exploring ergonomic rigging and joint amplification for precise hand ray selection in XR. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. ACM, Madeira, Portugal, 28–39. doi:10.1145/3715336.373567

[42] Matteo Marucci, Gianluca Di Flumeri, Gianluca Borghini, Nicolina Sciaraffa, Michele Scandola, Enea Francesco Pavone, Fabio Babiloni, Viviana Betti, and Pietro Aricò. 2021. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports* 11, 1 (01 Mar 2021), 4831. doi:10.1038/s41598-021-84196-8

[43] Jelena Mladenović, Jérémy Frey, Manon Bonnet-Save, Jérémie Mattout, and Fabien Lotte. 2017. The Impact of Flow in an EEG-based Brain Computer Interface. arXiv:1706.01728 [q-bio.NC] https://arxiv.org/abs/1706.01728

[44] Lennart E. Nacke and Craig A. Lindley. 2010. Affective Ludology, Flow and Immersion in a First- Person Shooter: Measurement of Player Experience. arXiv:1004.0248 [cs.HC] https://arxiv.org/abs/1004.0248

[45] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. The concept of flow. In *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. Springer, Sydney, Australia, 239–263. doi:10.1007/978-94-017-9088-8_16

[46] Sharon Oviatt. 1999. Ten myths of multimodal interaction. 74-81 pages. doi:10.1145/319382.319398

[47] Ricardo Pardo, Ignacio Miralles, Águeda Gómez-Cambronero, and Inmaculada Remolar. 2026. Analyzing the effectiveness and satisfaction of hand tracking vs. controllers among VR-experienced users. *Virtual Reality* 30, 2 (2026), 69. doi:10.1007/s10055-026-01333-2

[48] Jean-Éric Pelet, Saïd Ettis, and Kelly Cowart. 2017. Optimal experience of flow enhanced by telepresence: Evidence from social media use. *Information & Management* 54, 1 (2017), 115–128. doi:10.1016/j.im.2016.05.001

[49] Richard Rodriguez, Brian T Sullivan, Mayra Donaji Barrera Machuca, Anil Ufuk Batmaz, Cyane Tornatzky, and Francisco R Ortega. 2024. An artists' perspectives on natural interactions for virtual reality 3D sketching. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM SIGCHI, Honolulu, HI, USA, 1–20. doi:10.1145/3613904.3642758

[50] Anastasia Ruvimova, Junhyeok Kim, Thomas Fritz, Mark Hancock, and David C. Shepherd. 2020. "Transport Me Away": Fostering Flow in Open Offices through Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376724

[51] Florian Scholz, Patrick Rödig, Roman Rädle, et al. 2021. WoZ4U: An Open-Source Wizard-of-Oz Interface for Easy, Efficient WoZ Studies. In *Proceedings of the 2021 Conference on Human-Robot Interaction (workshop/demo)*. Frontiers Media SA, Lausanne, Switzerland, 668057. Open-access version available at PubMed Central. doi:10.3389/frobt.2021.668057

[52] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300590

[53] Smoothstep LLC. 2016. *Quill*. https://quill.art/

[54] Ivan E. Sutherland. 1963. Sketchpad: a man-machine graphical communication system. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference* (Detroit, Michigan) *(AFIPS '63 (Spring))*. Association for Computing Machinery, New York, NY, USA, 329–346. doi:10.1145/1461551.1461591

[55] Penny Sweetser, Daniel Johnson, and Jay Kyburz. 2020. Evaluating GameFlow in a Multiplayer Online Strategy Game Under Development. In *Proceedings of the Australasian Computer Science Week Multiconference* (Melbourne, VIC, Australia) *(ACSW '20)*. Association for Computing Machinery, New York, NY, USA, Article 47, 10 pages. doi:10.1145/3373017.3373068

[56] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Comput. Entertain.* 3, 3 (July 2005), 3. doi:10.1145/1077246.1077253

[57] Gershon Tenenbaum, Gerald J Fogarty, and Susan A Jackson. 1999. The Flow Experience: A Rasch Analysis of Jackson's Flow State Scale. *Journal of outcome measurement (Wheaton, Ill.)* 3, 3 (1999), 278–. https://pubmed.ncbi.nlm.nih.gov/10431494/

[58] B. C. G. van Houwelingen, A.-F. Rutkowski, S. Ganni, P. S. Stepaniak, and J. J. Jakimowicz. 2020. Effects of surgical flow disruptions on surgeons' resources: a pilot study. *Surgical Endoscopy* 34, 10 (01 Oct 2020), 4525–4535. doi:10.1007/s00464-019-07239-2

[59] Shan Wang and Shani Liu. 2025. Design and Implementation of a Multi-Sensory Interactive Virtual-Reality Platform for Artistic Creation. In *2025 5th International Symposium on Computer Technology and Information Science (ISCTIS)*. IEEE, IEEE, Xi'an, China, 1044–1047. doi:10.1109/ISCTIS65944.2025.11066040

[60] Adam S. Williams, Jason Garcia, and Francisco Ortega. 2020. Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3479–3489. doi:10.1109/TVCG.2020.3023566

[61] Bob G Witmer, Christian J Jerome, and Michael J Singer. 2005. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments* 14, 3 (2005), 298–312. doi:10.1162/105474605323384654

[62] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM SIGCHI, Vancouver BC Canada, 143–146. doi:10.1145/1978942.1978963

[63] Anna Zhilyaeva. 2018. Virtual reality live performance at the Louvre Museum Paris. https://www.annadreambrush.com/tiltbrush/virtual-reality-live-performance-louvre-museum-paris/. Blog post.

[64] Xiaoyan Zhou, Adam Sinclair Williams, and Francisco Raul Ortega. 2022. Eliciting Multimodal Gesture+Speech Interactions in a Multi-Object Augmented Reality Environment. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22)*. ACM, Tsukuba, Japan, 1–10. doi:10.1145/3562939.3565637