TouchPilot: Designing a Guidance System that Assists Blind People in Learning Complex 3D Structures

Xiyue Wang
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
wang.xiyue@lab.miraikan.jst.go.jp

Seita Kayukawa
IBM Research – Tokyo
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
Seita.Kayukawa@ibm.com

Hironobu Takagi
IBM Research – Tokyo
Tokyo, Japan
takagih@jp.ibm.com

Chieko Asakawa
Miraikan – The National Museum of Emerging Science and Innovation
Tokyo, Japan
chiekoa@us.ibm.com

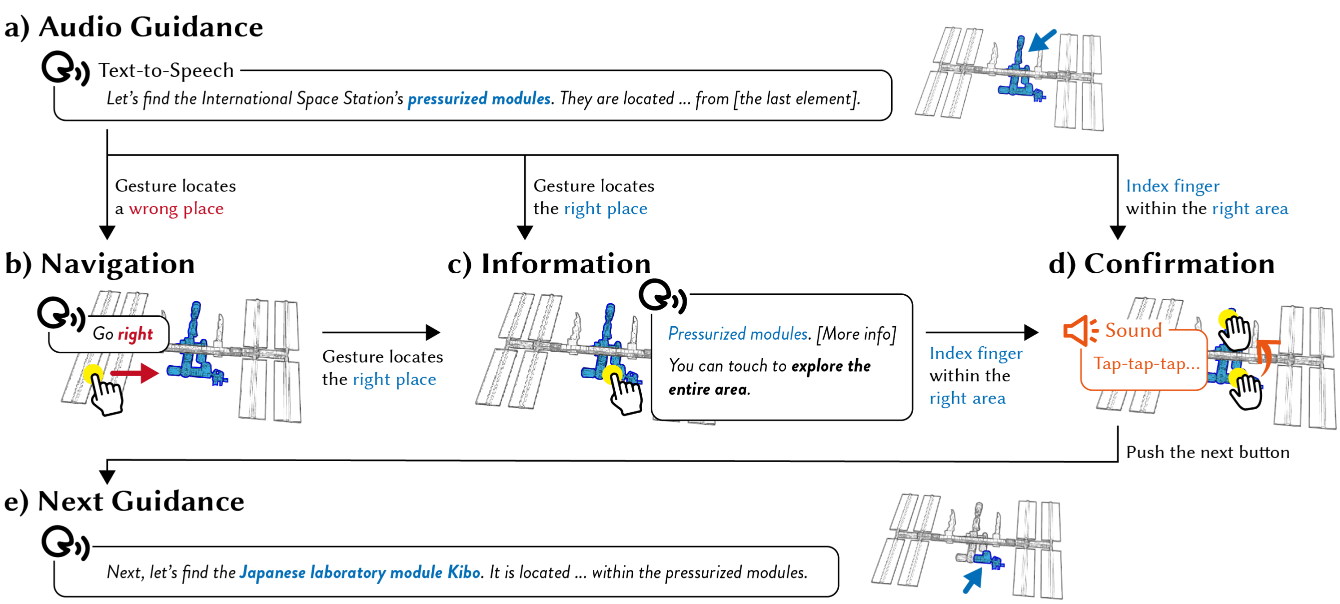


Figure 1: Flow of hand interaction in TouchPilot's guidance system, which supports step-by-step learning of the textual and spatial information. a) TouchPilot provides audio guidance of a target element via text-to-speech. b) If the user points to the wrong place, the system provides direction guidance. c) When the user points to the correct place, the system announces its information. d) The system encourages the user to explore the entire area while listening to the confirmation sound (the "tap-tap-tap" sound). e) The system provides the next guidance after the user pushes the next button.

Author Version

ABSTRACT

Making complex structures accessible to blind people is challenging due to the need for skilled explainers. Interactive 3D printed models (I3Ms) have been developed to enable independent learning of 3D models through activating audio labels. However, they present single-layered information and require users to identify interactive elements through a pinpointing action, which might be insufficient for learning complex and unfamiliar subjects. In this paper, we investigate I3Ms for complex structures. We propose TouchPilot, a guidance system designed based on a study that observed learner-explainer interaction styles. TouchPilot guides users step by step through navigation, exploration of hierarchical elements, and confirmation of their entire areas. A follow-up study found that the guidance system led to better learning outcomes and higher independence compared to a pinpointing system. Feedback suggests that being primed by the guidance system systematically, followed by pinpointing freely for review, is preferred for learning complex structures.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**; *Interaction techniques.*

KEYWORDS

interactive 3D printed models, visual impairments, guidance, computer vision

ACM Reference Format:
Xiyue Wang, Seita Kayukawa, Hironobu Takagi, and Chieko Asakawa. 2023. TouchPilot: Designing a Guidance System that Assists Blind People in Learning Complex 3D Structures. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23), October 22–25, 2023, New York, NY, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3597638.3608426

1 INTRODUCTION

The increased availability of high-fidelity 3D printers [7, 25, 68] and open data of 3D models [3, 6] has enabled the visually impaired to gain access to a wide range of tactile objects. For example, 3D printed models have been used to teach the structure of cells, molecules, machines, and celestial bodies in studying abstract science concepts [27, 46, 58, 63]. When a structure becomes complex, explanations are needed. The structure can be considered complex if the user does not have previous experience with it or something similar [12], if it contains varied and multiple-layered components [43], or if it has advanced functionality such as moving or removable parts [45]. Making such complex structures accessible and easy to understand for the visually impaired is an important goal for STEM education [29, 58, 63] and for museums that provide universal access to the public [44, 66, 67].

Research on interactive 3D printed models (I3Ms) has aimed to help visually impaired users learn structures independently through audio explanations during tactile exploration. However, a key characteristic of current systems is that they rely primarily on users adopting a pinpointing style to identify specific elements and activate the audio labels [9, 10, 43, 47, 48, 51]. The pinpointing-based systems could be challenging for blind users in their efforts to identify crucial information and capture the overall picture of a structure with many audio labels. Moreover, most I3Ms have been limited to simple structures by allowing only a few non-overlapping elements in pinpointing interaction. Such systems could not adequately represent the varied and multi-layered information of complex structures. Conversely, hierarchical information has been provided for exploring 2D images and tactile reliefs [30, 43], but such systems could still encounter challenges in finding information, especially in the subdivided regions [43]. Furthermore, none of them assist in capturing the shapes of complex 3D structures. To the best of our knowledge, there has been no investigation of I3Ms for assisting comprehensive and independent learning of complex 3D structures. To fill this research gap, we investigated the following research question in three stages: *What kind of I3Ms can assist blind individuals in learning complex structures independently?*

In our first study, Study 1, we conducted an open-ended investigation of six paired sessions between participants and explainers to examine natural interaction styles when learning complex models. Through analysis of communication styles, we observed that blind participants did not actively acquire information, despite previous I3M systems requiring active information acquisition. Our findings suggest that guidance is necessary for learning such complex structures. This guidance involved proactive introduction of hierarchical information, as well as assistance in navigation and identification of specific elements.

Based on Study 1's findings, we developed *TouchPilot*, a novel step-by-step guidance system that guides users through exploring the textual and spatial information of hierarchical elements on a complex model using optical hand tracking (Fig. 1). At each step, users could point to the target element using gestures, and receive direction guidance if they pointed to the wrong place (Fig. 1b). When pointed to the correct place, the system announced its information (Fig. 1c). Moreover, users could touch the element with their full hand to explore its entire area and shape, with a confirmation sound activated when the index finger entered the correct area (Fig. 1d). The system also allowed users to control the steps and pace of their learning (Fig. 1e).

We implemented the previously developed pinpointing-based system as a baseline and conducted Study 2 to compare the two systems in terms of knowledge-based questions and self-ratings. Eight blind participants with no prior knowledge of the models participated in this study. The results show that (1) the guidance system led to better learning outcomes for understanding hierarchical textual information and spatial information, (2) blind users were more confident that by using the guide system, they could learn such complex structures independently, and (3) most participants found the guidance system easier to use. Moreover, all functions of the guidance system were found to be useful for understanding the subject. Regarding real-world usage, all participants preferred using the guidance system initially to prime themselves with structured knowledge, followed by using the pinpointing system to review the information based on their interests. We discuss future possibilities for integrating the guidance system with pinpointing and other customized inquiry-based systems to support systematic and free explorations of complex structures.

2 RELATED WORK

2.1 Accessible Tactile Objects and Models

For people with vision impairment, touch is one of the most important modalities for learning about the world. Researchers have found that using real objects is preferred in educational settings to represent the actual scale and tactile sensations [36, 57, 59]. However, they are not always practical when too large, small, or dangerous. As an alternative, the growing prevalence of fabrication technologies has made replicas available in tactile graphics [1, 34, 70] and 3D models [2, 5, 45, 48, 68].

Compared to tactile graphics, which give a relatively flat representation of visual information using raised lines and areas [23], commodity 3D printing offers unbounded possibilities for representing three-dimensional concepts and information. Literature has found that 3D models are preferred in orientation and mobility training [19, 20, 23, 60, 69] and STEM education [27, 29, 37, 46, 58, 63] as they can give learners a sense of spatial awareness [14, 69]. One of the difficulties in creating 3D models is that it requires considerable time and technical skills to design the model [8, 15, 36]. Nowadays, 3D scanning has made it possible to reproduce scientific and historical artifacts that are hard to model [11, 32, 33, 56]. Moreover, open 3D datasets have become available [3, 6], which promotes the accessibility community to produce tactile objects for people with vision impairment. The complex 3D printed models have also been designed with varied complexity [58] and removable parts [45] for easy understanding. Guidelines have been proposed to design and adjust 3D models for better touch perception [23, 24, 58].

2.2 Interactive 3D Printed Models (I3Ms)

Tactile objects often feature braille labels, but they are limited by space [32, 45, 49], and not all visually impaired individuals can read braille [35]. Research has explored creating interactive 3D printed models (I3Ms) with audio labels to augment tactile experiences and enhance engagement [23, 54, 61]. Compared to braille labels, audio labels on tactile objects have been shown to significantly improve map reading efficiency and satisfaction [4] and space and text memorization in STEM education [14]. The triggering of the audio label should be separated from tactile examination actions to avoid confusion resulting from unintentional audio (e.g., the Midas Touch effect) [26, 43, 50].

Various methods have implemented physical audio labels on tactile objects, including push-buttons [23, 31, 61], touch screens [17, 62], acoustic sensing [49, 55], and NFC tags [11]. However, a major limitation of these implementations is that the physical audio labels must be designed on and with tactile objects, which limits the number of labels that can be implemented and creates challenges for complex 3D models.

Optical-based hand and 3D object recognition have opened up new possibilities in terms of the unconstrained audio label shape and location and various gestures to trigger audio. Wilson introduced using a depth camera to detect a touch on an uninstrumented surface [65]. Access Lens[28], Tactile Graphics Helper[13], and Molder[52] then implemented camera-based hand tracking to make documents, tactile graphics, and maps accessible through triggered audio where the finger points. Optical hand tracking was further expanded to enable audio-tactile exploration of tactile reliefs. Buonamici et al. used a Microsoft Kinect to trigger audio when the right index fingertip points to a particular area [9], and Reichinger et al. used Intel RealSense to trigger audio with gestures [42, 43]. CamIO extended the concept to touch interaction on 3D objects with at least two different audio labels [47] and investigated the annotation of labels [10]. Shi et al. developed Markit and Talkit [51], a low-barrier toolkit for creating and interacting with audio annotations on 3D models, and adapted them to classrooms to investigate I3Ms design guidelines for model and audio guide [48].

However, prevalent I3Ms require visually impaired users to pinpoint spatially to identify specific elements and acquire information. The effectiveness was not examined with complex structures that are unfamiliar, have a lot of elements, or advanced functions. Moreover, the pinpointing-based I3Ms struggle to distinguish between hierarchical elements, as most of them implemented a single-layer information structure with a few non-overlapping audio labels. As a result, the pinpointing-based I3Ms might not be sufficient for learning 3D structures that are complex.

2.3 Guidance for Tactile Exploration

Identifying an object becomes difficult when one does not have previous experience with it or something similar [12]. Blind individuals often require assistance exploring and locating information. In appreciating art, skilled guide people often held the hands of the visually impaired in exploring 3D reproductions from paintings [9, 22]. Similarly, in learning complex science concepts and structures, sighted learners assisted the visually impaired during exploration following a lecture on a 3D telescope model [58].

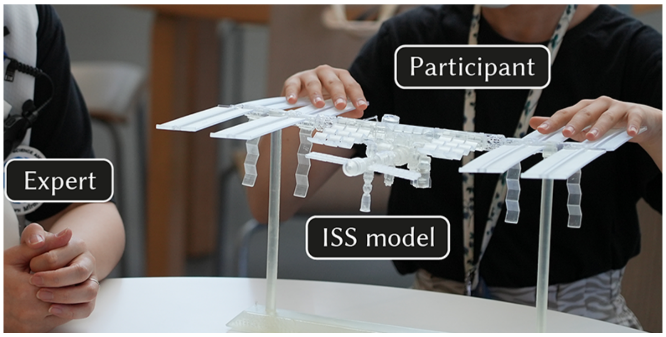
However, the guide received from others is often perceived as a limitation for the blind who increasingly demand autonomous accessibility [9]. To address this, assistive technologies have been developed to guide blind individuals to locate tactile objects independently. Vibration cues have been developed on touchscreen [16] and smartwatch [18] to enable navigation and tracing of graphical information. Tangible and movable markers [53] and miniature robot [38] have been proposed to guide the hand physically to targets. Ramoa et al. compared three types of audio navigation cues for blind users: sonar, axis, and voice, and found that the voice method was the fastest without prior training [40]. Guo et al. developed VizLens, which gives interactive feedback and guidance for learning an interface, and importantly, they found that users preferred guidance when using unfamiliar interfaces [21]. However, their system required users to know and preselect the desired target to provide such guidance.

In addition to guiding locations, researchers have explored supporting blind individuals in understanding complex information using information hierarchies. For 2D images, Lee et al. developed ImageExplorer, a multi-layered image exploration system that enables users to explore the spatial layout and information hierarchies of images [30]. Reichinger et al. implemented hierarchical exploration for tactile reliefs, with audio labels of subdivided parts only becoming available once all six basic regions have been explored [43]. However, the outcome of the hierarchical exploration was not evaluated, and they observed some participants did not reach the final layer.

Despite these efforts in guiding 2D explorations, guidance for tactile exploration of complex 3D structures remains underexplored. We fill the research gap by designing a guidance system targeting learning complex 3D structures.

3 STUDY 1: INITIAL EXPLORATION OF INTERACTION STYLES

Our first study aimed to gain insights into the natural interaction styles in learning a complex structure through tactile exploration. We constructed one-on-one sessions where each blind learner was



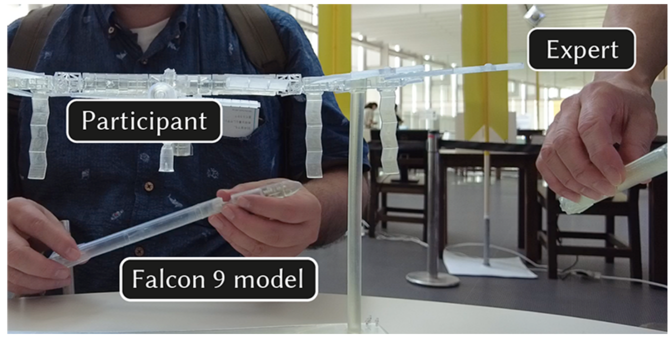


Figure 2: Participants learn the ISS and Falcon 9 models, accompanied by one expert.

accompanied by an expert to give explanations when necessary. We also collected feedback from learners and experts to better understand the current interaction system's strengths and challenges for potential improvements. Based on our findings, we discussed the limitations of the pinpointing-based I3Ms, and we derived design guidelines for I3Ms that could better support users in learning complex structures.

3.1 Participants

The study was conducted as part of an event at a science museum for the visually impaired. We recruited six blind participants (three females) aged 13 to 64 years (mean = 48.2, SD = 20.4) and three experts who specialize in communicating space science to museum visitors. All experts received training in how to explain exhibits to visually impaired visitors. One had experience in designing and teaching with 3D models for the visually impaired after receiving guidance from teachers at a blind school. The experts assisted with the content design in advance and also served as explainers.

3.2 Apparatus

In collaboration with three experts, we developed two 3D printed models related to pioneering space technologies. This topic was chosen based on the science museum's activities to promote accessibility. Despite frequent media coverage of human achievements in space, its fundamental knowledge was unfamiliar to most visually impaired individuals. On the other hand, the museum's previous experiences demonstrated that it was one of the most highly anticipated topics for visually impaired visitors. Notably, the topic is complex, and the textual guide was insufficient for blind visitors to mentally construct a structural image. Accordingly, we aimed to make the topic more accessible using 3D printed models.

Two models were sourced from Thingiverse.com and modified to suit tactile exploration. Each contains more than ten different elements, and each element belongs to higher layer structures. Some elements are removable parts connected with magnets, allowing the models to enable users to experience engineering functions tactically and dynamically, such as space shuttle docking and rocket launching. These characteristics fulfill the requirements of a complex structure.

- International Space Station (ISS, Thing: 4203169[1]), the largest manmade structure in space. We thickened the main solar and radiator panels to make them more robust for tactile examination. Trivial antennas, robot arms, and equipment were removed to facilitate tactile examination of the essential parts. A magnet was also embedded to support docking of a space shuttle model.
- Falcon 9 (Thing:4503875[2]), an innovative and reusable rocket that carries a crew and cargo to the ISS. We integrated the original multi-part data into four main parts: Stage 1, Stage 2, the trunk part of the Crew Dragon space shuttle, and the capsule part of the shuttle. The match-and-rotation joints between Stages 1 and 2 were reinforced to support easy separation. Magnets were attached to the contact surfaces between Stage 2, the trunk, and the capsule for easy separation and connection.

The models were printed at a 1/267 scale of the actual objects to support general exploration and detailed examination. A stand was designed to hold the ISS in place for two-hand exploration.

3.3 Procedure

Each participant was accompanied by an expert and explored both models (Fig. 2). To reflect the inquiry style used for most I3Ms, participants were instructed to initiate conversations by asking questions during the exploration. The experts introduced each session by stating, "This is a model of... in front of you. You can touch it freely and ask me questions anytime."[3] Meanwhile, to reflect natural communication styles, experts were instructed that they could initiate interactions if participants were inactive and they felt an explanation was necessary. Each session lasted for a maximum of 30 minutes, and participants could end it early. Consequently, we gathered feedback from the participants regarding the strengths of the system and areas that could be improved. Experts also provided insight into places that were difficult to explain.

3.4 Analysis and Label Categorization

We video-recorded the sessions. Two researchers developed labels and themes from the recording transcripts using open coding and

[1]https://www.thingiverse.com/thing:4203169
[2]https://www.thingiverse.com/thing:4503875
[3]All communication with the participants was in their native language. In this paper, we present any translated content in the form of *"translated content."*

Table 1: The 18 categorized activity labels, their contents, and example words. All examples were collected from actual sessions.

From the Experts		
Label	Content	Examples
E: General	General information	<1> “There is a model of the ISS on a stand in front of you. It is about 400 km from the ground, orbiting Earth about once every 90 minutes.”
E: Basic	Basic element	<2> “This a solar panel. [Extension: It generates the electricity that the ISS uses.]” <3> “This part is called Soyuz. [Extension: It is a spacecraft made by Russia. Now it is docked to the ISS but it can be separated...]”
E: Composite	Composite of several basic elements	<4> “This large area contains all modules. [Extension: These parts are where the astronauts live.]” <5> “This part has modules made by Russia’s space agency. [Extension: The shape of these modules is different from the others. They are more slender and irregular.]”
E: Context	Contextual information	<6> “Actually, there are two names for astronauts. In North America, Europe and Japan, people call them ‘Astronauts.’ But in Russia, people call them ‘Cosmonauts.’ ”
E: Extension	More information	See extension parts of examples <2> <3> <4> <5>.
E: Question	Raise a question	<7> “Do you want to travel to space someday?”
E: Navigation	Guide the hand to a place	<8> “”"If you move your right hand up slightly, you can find something jagged. [Basic: This is. . .]" <9> “Can I touch your hand?” (If approval given, move the hand.) “[Basic: From here to here is. . .]”
E: Specification	Specify the place to be explained	<10> “Your left-hand fingers are now touching something. [Basic: This is ...]”

From the Participants		
Label	Content	Examples
P: General-IN	An initial question about general information	No participants performed this action.
P: General-FL	A follow-up question about general information	No participants performed this action.
P: Basic-IN	An initial question about the basic element	<11> “What is this zigzag thing over here?” (With hand movement to indicate the area)
P: Basic-FL	A follow-up question about the basic element	<12> (After hearing <2>) “(The solar panels) Are they unfolded during the space flight? How were they transported to space?”"
P: Composite-IN	An initial question about a composite	<13> “People go into these places?” (With hand movement to indicate the area)
P: Composite-FL	A follow-up question about a composite	<14> (After hearing <4>) “So, can astronauts move freely across these modules?”
P: Context-IN	An initial question about contextual information	<15> “Is space junk a problem right now?”
P: Context-FL	A follow-up question about contextual information	<16> (After hearing <6>) “Do they have different kinds of qualifications?”
P: Confirmation	Verbally confirm the location	<17> “You mean here, right?” (And pointing.) <18> “(I found the place.) It is flat, like a roof.”
P: Examination	Examine the model in silence	No word has been said.

axial coding [64]. We categorized activities by observing the sessions, considering the previous experience of designing I3Ms. Three types of activities were initiated by both participants and experts:

- General information: An introduction of the model, including its overview and scale.
- On-model information: Explaining part of the model. We further divide the information into basic elements and composites when it is hierarchical. Basic elements refer to the

rudimentary areas, and composites are areas composed of several elements. If the information is single-layered and thus does not contain a hierarchical structure, we categorize all information into basic elements.
- Contextual Information: Explaining the background story and facts related to the subject.

To investigate how communication was constructed, we specified who initiated the activity. The activity was an explanation if initiated by the expert and a Q&A if initiated by the participants. We further categorized Q&A into initial and follow-up questions. Initial questions were proposed by participants without any prior knowledge explained, while follow-up questions were additional ones asked after hearing the expert's explanations.

In addition to these categories, we identified unique activities performed by experts and participants. The expert-specific activities include:

- Extension: Providing more information about a basic element or a composite.
- Question: Raising a question with the participant to increase engagement.
- Navigation: Verbally guiding a participant's hand to a specific location, or physically holding their hand to move to a location.
- Specification: Specifying the place the explanation refers to when the participant examines the model with both hands.

The participant-specific activities include:

- Confirmation: Verbally confirming understanding after hearing the navigation or specification guidance from experts.
- Examination: Tactilely exploring the model without interacting with the expert.

A total of 18 activities are categorized, eight from the experts and ten from the participants. The content and examples of each label are presented in Table 1.

3.5 Results and Findings

We analyzed the count of each activity to reflect its frequency as the duration of each activity varied. Overall, the ISS sessions had more activity counts (mean = 33.50, SD = 22.41) than the Falcon rocket sessions (mean = 10.34, SD = 6.58), but similar trends were discovered. Our analysis provided insights into the styles and trends of expert-participant interactions.

3.5.1 Expert Guided the Sessions. We first compared the counts of explanations initiated by experts, which include E: General, E: Basic, E: Composite, and E: Context labels, and inquiries initiated by participants, which include P: General, P: Basic, P: Composite, and P: Context labels, grouped by Initial and Follow-Up types as shown in Fig. 3. It was found that the number of expert-initiated activities exceeded that of the participant, except for the E2P2 (ISS, Falcon 9) and E1P5 (Falcon 9) sessions. Within participant-initiated activities, initial questions were found to be less common than follow-up ones in all sessions. Moreover, except for P2, none of the participants asked initial questions about the Falcon 9 model (Fig. 3b). This finding indicates that it might be easier for participants to inquire after receiving information from the experts. Even though we encouraged the participants to make inquiries (as described in Section 3.3), the sessions still leaned toward experts giving information rather than participants actively seeking information, since the participants did not frequently ask initial questions and the experts felt the need to initiate explanations.

3.5.2 Hierarchical Information. Next, we normalized the counts into the percentage within the session and investigated the mean percentage of each activity across all sessions, as summarized in Fig. 4. We found that both expert explanations and participant inquiries included basic elements and composites. The explanations for basic elements (E: Basic) and composites (E: Composite) had the highest percentages among all explanations (E: General, E: Basic, E: Composite, and E: Context). For the ISS model, the percentage of expert explanations for composites was slightly less than for basic elements, but this was reversed for the Falcon 9 model. This implies that hierarchical information was common in these complex structures. For example, experts explained the ISS model by first introducing the entire module area and then going deeper into different modules. For the rocket model, when experts introduced topics such as its launch, they explained structures from the general components (different stages) to the details (engines at each stage.) These explanations were usually constructed in steps, in a top-down manner from composites to basic elements, as described in the above examples. We observed that an ideal complexity level for public learning of such models involves two to three general topics with one to two layers of hierarchical information.

We also found that extension was the most common activity for both models (Fig. 4), supporting the need for "more information," as previously noted in the I3M literature [50, 51].

3.5.3 Navigation and Confirmation. Navigation and specification guides were also found in sessions with a high percentage (Fig. 4). The experts utilized navigation to assist the participant in locating an element before explaining it systematically. The specification guide assisted the participant when further location confirmation was needed. These guides also included shape descriptions to help participants locate the desired element quickly. Participants also appeared to confirm the location during the sessions for both models (Fig. 4). After locating an element, they tended to reconfirm or react by describing the shape of the entire area to ensure they had located the correct place. After that, the expert explained as the participant explored the referred area.

3.5.4 Uncontrollable Pace. All participants praised the combination of models and explanations, stating that they were able to grasp the overall image more effectively with this format. Despite the fact that five out of six participants did not take the lead by actively initiating questions, none expressed dissatisfaction with the current level of autonomy and control over the session. Several participants (P4 and P6) expressed their preference to follow the guidance of the expert before attempting to ask questions, confirming the need for the guidance. However, one expert expressed that the biggest obstacle to a smooth session was the difficulty of determining the appropriate timing of explanations. Since examination without communication (P: Examination) appeared to be the third most frequent participant activity (Fig. 4), an interactive system that gives the participants independence in controlling the pace is necessary.

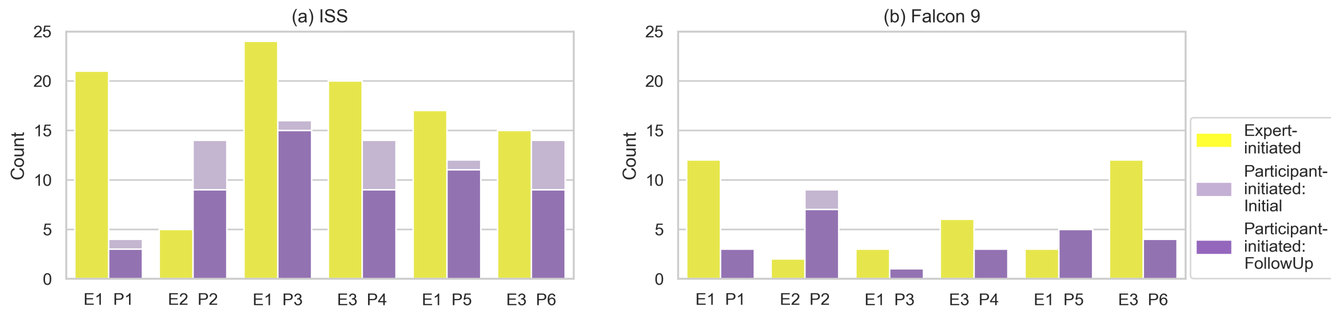


Figure 3: Overview of each session's expert-initiated explanation and participant-initiated inquiry counts. Each session is divided into two stacked bars. The left bar represents the count of activities initiated by the experts, while the right bar represents the count of activities initiated by the participants.

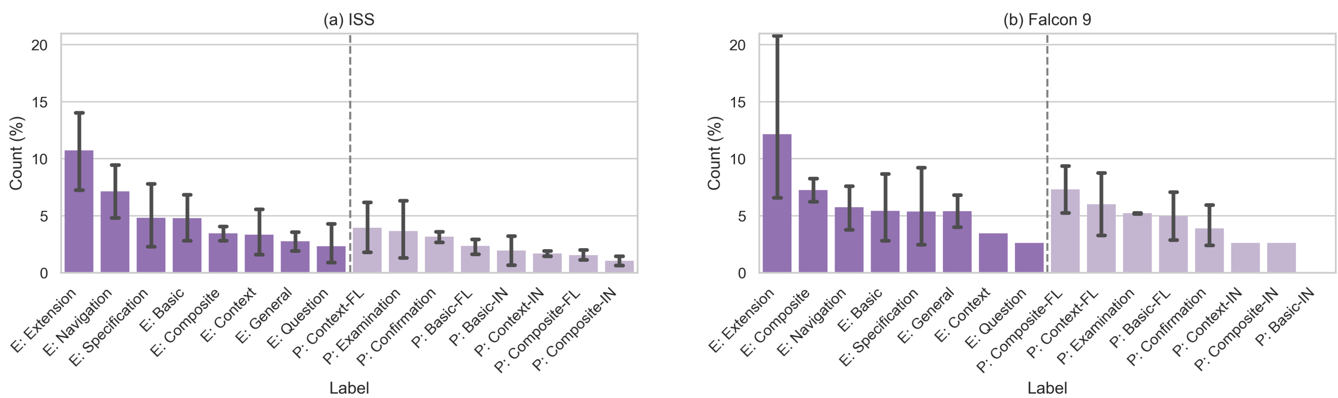


Figure 4: Mean percentage of activities. The percentage is calculated by dividing the activity count by the entire session count. The labels for expert-initiated activities and participant-initiated activities were separated into two groups, and activities within each group were arranged in descending order by the mean value.

3.5.5 Design Goals. The above findings suggest that the interaction with pinpointing- and inquiry-based I3M systems was insufficient to learn complex models. A new interactive system that guides users to acquire knowledge through tactile exploration is necessary. The system should include the following functions:

- Explanation of hierarchical information with the user controlling the pace.
- Navigation support to help users locate elements referenced in the explanation.
- Confirmation support to ensure that users have located the correct element and to help users learn its area.

The design goals encompassed fundamental needs found in natural interaction, although Context and Question from experts were unfocused due to their infrequent occurrence and the need for additional conversational functionalities. In addition, distinct gestures and controls that are separate from regular tactile examination should be designed to trigger the above guidance functions, following I3Ms practices [26, 43, 50].

4 SYSTEM DESIGN

We designed the guidance system for I3Ms based on our design goals and findings from Study 1. The underlying structure is a base system that transforms physical 3D models into point cloud data using 3D data processing and detects hand locations and gestures using computer vision. Two interactive systems were built on top of the base system for comparison: a pinpointing system, inspired by the previous I3M design [9, 43, 51], that triggers audio information by pointing and a guidance system that provides hierarchical explanations, navigation, and confirmation support. They operated in real-time without any latency observed later by users.

4.1 Base System

The base system contained a virtual representation of hands and 3D models for interactions. To collect 3D data, we implemented an optical solution using a depth camera (or RGB-D camera) and computer vision technology. The depth information detected touches on an uninstrumented surface by detecting how far the fingers are from the camera [43, 65]. The computer vision-based solution allows areas on a 3D model to be annotated and retrieved freely without extra electronics [51]. In our setup, a 4×5 checkerboard

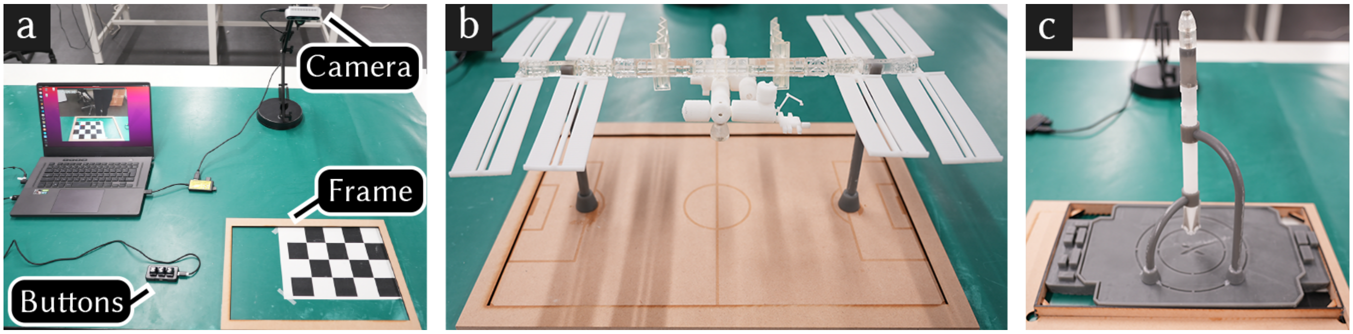


Figure 5: From left to right: system setup (a), ISS model (b) and Falcon 9 model (c), both are reinforced for steady-hand exploration.

pattern with $5cm$ squares was placed on a flat surface to mark the world coordinate. A Realsense D435 depth camera was positioned about $45cm$ away and $45cm$ above the origin at a $35degree$ angle pointing downward (Fig. 5a). The camera's field of view covered the checkerboard, the 3D model, and the hands examining the models.

The system did not support real-time recalibration, so the 3D models were adjusted to stay firmly in place. We created sturdy stands and bases for both models: a laser-cut soccer field at the same scale as the ISS (Fig. 5b) and a cargo ship used during Falcon 9 Stage 1 recovery and recycling (Fig. 5c). Additionally, we made the landing leg of the Falcon 9 model movable to make the recycling narrative more interactive. The 3D model data were imported as a point cloud using the open-source library Open3D [72]. We built a custom annotator similar to one described previously [51]. Designers could select points representing an element, annotate them with a tag, and export the data into a JSON file. We then imported the base model and annotated data into the world coordinate system of the virtual space, representing the physical model's location in real space (Fig. 6). For a movable element, the base system tracked its initial position, and audio instructions were provided in both pinpointing and guidance systems for movement and recovery.

Hand detection was achieved using MediaPipe Hands [71], a real-time machine learning solution that detects and tracks 21 landmarks of each hand on RGB frames. We converted the detected 2D landmarks into 3D landmarks using depth values and mapped them from the camera coordinate system into the world coordinate system.

4.2 Pinpointing System

A gesture-controlled system was implemented to distinguish audio elicitation commands from regular tactical examinations. Following prior studies on audio label elicitation [9, 43, 51], we designated the index finger as the finger that identifies an element, and the "number one" gesture as the command to activate the audio label. The "number one" gesture involves pointing the index finger upward while keeping the other fingers in a fist. We confined the activation control to one hand (default: right) and a non-moving gesture (for $0.2s$) to minimize false activation and confusion of mismatched information and location, based on previous gesture studies [43, 50]. However, users can customize their preferred hand, finger, and gesture for operating the system.

The "number one" gesture was detected using a deep neural network (DNN) trained on 3D landmarks. Training and testing data were collected from three individuals with different hand sizes (one female). They were asked to perform the "number one" gesture to touch the entire model for 1 minute, followed by tactile exploration with both hands for 1 minute to collect positive and negative data. The classifier identified the gesture with an acceptable precision score (0.82), recall score (0.98), and F1 score (0.89) with no sign of over-fitting.

To identify a touched element, we generated a $5mm$-radius sphere at the tip of the index finger and detected 3D model points within the sphere (Fig. 6). If multiple annotated elements were detected, we identified the touched element as the one with the largest contact area. Since the element being touched could not be distinguished if one element was composed of others, no hierarchical element should be added to the system.

Seventeen non-overlapping basic elements were annotated for the ISS model and eleven for Falcon 9. The system used a CSV file to store the textual data with three headers: Tag, Name, and Content. It used Watson Text-to-Speech to generate a voiceover. The Content was announced immediately after the Name since Study 1 found it natural to hear extensions after the basic explanation. The text for each voiceover was kept under 50 words to avoid being overwhelming. A button was implemented to stop the audio (Fig. 6), and a beep sound was played when the explanation ended.

4.3 Guidance System

The guidance system used the same gesture as the pinpointing system to activate audio interactions but also included several types of interactions (Fig. 1) to introduce elements step by step (Fig. 7). Therefore, designing and implementing a guidance system that provides these additional features can be more challenging than implementing a pinpointing system.

4.3.1 Interaction Flow. The following steps were designed to provide a hierarchical exploration of a model:

(1) The system introduces the model's name, scale, and general information.
(2) The first layer of elements is introduced with an overview, followed by the name and location of the first element. The user is asked to locate the target element.

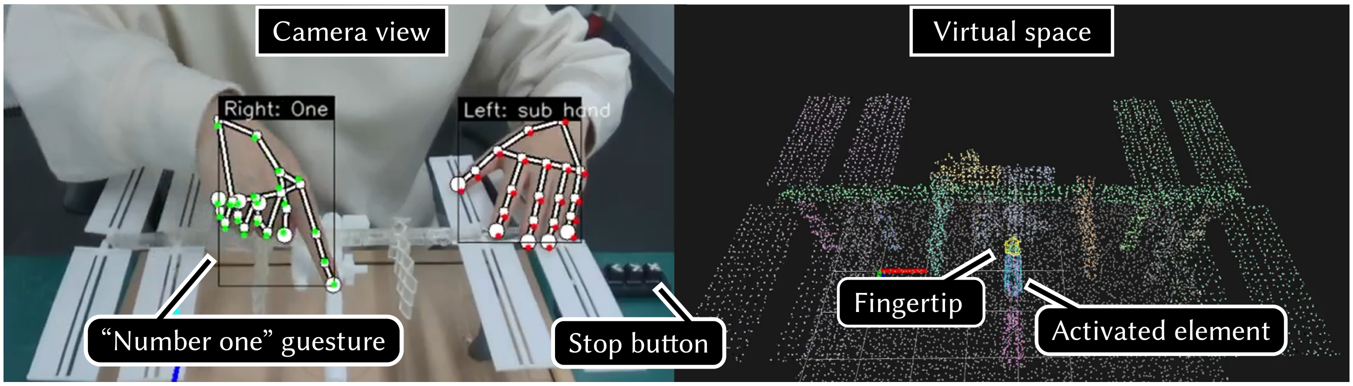


Figure 6: Pinpointing system with the ISS model. Users activate the audio label by using "number one" gesture. The system detects the index fingertip position (marked with a yellow sphere in the virtual space) and identifies the element in contact with it (marked with a blue wireframe.) All elements (point clouds marked with different non-grayscale colors) are available for acquiring information.

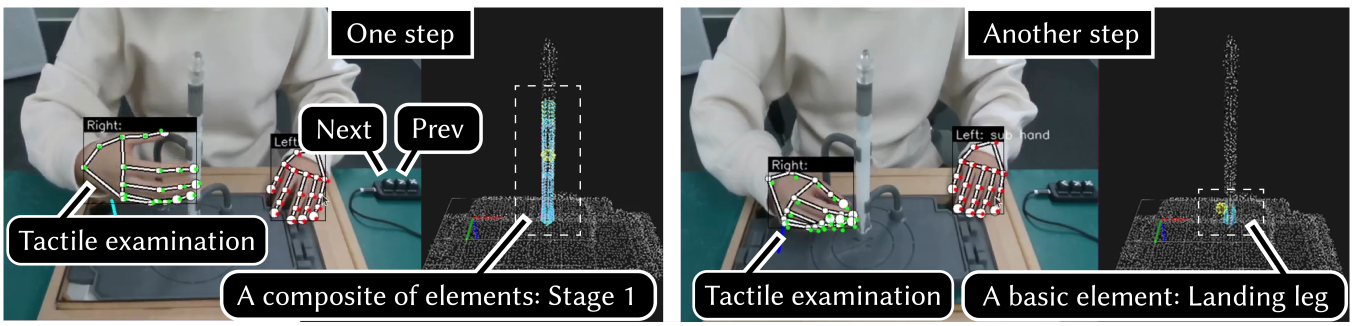


Figure 7: Guidance system with the Falcon 9 model. One element is available for interaction at each step. The user is confirming the entire area using the regular tactile examination. The available element can be a composite of elements or a basic element (both marked with a blue wireframe), and other elements are unavailable for interaction (point cloud marked in grey.)

(3) Navigation and confirmation guides are available for spatial exploration. If the user points to the wrong area using the "number one" gesture, the navigation guide is provided. If the user points to the correct area, an explanation is given. During the normal examination, a tapping sound plays for confirmation when the index finger makes contact with the correct area, and no sound plays otherwise.
(4) After the element is located and the explanation is heard, optionally, the system encourages the user to explore the entire area while listening to a distinct confirmation sound.
(5) The user presses the Next button to access the next element of the block, which is introduced with its name and relative location from the previous element. The user is then asked to locate it.
(6) When all elements in a layer are examined, the next layer becomes available, and its overview is announced. Then the user is asked to locate all elements in this block in order.

This flow was designed to provide a structured approach to guide the building of hierarchical and spatial knowledge. It first introduces a composite of several elements and then gradually moves to the individual basic elements, providing guidance and confirmation at each step.

4.3.2 Function Design. The hierarchical textual data is structured in a CSV spreadsheet. In addition to Tag, Name, and Content used in the pinpointing system, each line of data includes a Block, Location, and Encourage Shape flag. The layered structure is achieved using blocks. Each block is a group of elements associated with one sub-scenario. The Block value is formatted as "X_Y" where X represents the order of the entire block and Y represents the element order within the block. This data structure allows the system first to introduce the overview of the group, then dive into the elements one by one systematically, even if they overlap. Spatial information, which indicates the spatial relationship between elements, is provided through Location annotation, in which the annotator specifies the element's location within the model. This annotation is optional, and if not given, the system automatically calculates the location based on the direction from the previous element. Setting Encourage Shape to true produces an additional sentence ("You can explore its entire area by touch.") after the explanation to encourage users to explore the entire shape tactilely. The ISS includes five blocks (major components, the pressurized modules of each country, Japanese modules, airlocks for extravehicular activities, and spacecraft docking) and a total of twenty steps. The Falcon 9 model includes five blocks (company introduction, major components, spacecraft components, launching, and recovering) and twenty steps.

The navigation function implements the KDTree algorithm[4] to calculate the distance between the fingertip and the closest point in the target element. To aid navigation, our system provides direct voice instructions such as "go up" or "go down," which were more effective than sonar or axis-based sounds in previous studies [41]. To simplify the precise distance in all three axes, our system reports only the axis with the maximum distance. If the second-highest distance exceeds $2cm$, we include the second axis, for example, "go left, then go down." Additionally, if the target is within $1cm$ of the fingertip, we add "a little" to the direction, such as "go a little left."

The confirmation feature is designed to help the user quickly identify the correct element and capture the entire shape during normal examination. It can be activated without using the pointing gesture. When the contact area of the user's index finger is within the area of the target element, a tapping sound at a one-second interval is activated.

In addition, the system includes buttons for moving the guidance to the next and previous elements (Fig. 7), as well as sound cues indicating whether the operating hand is being tracked. When the user's hand is within the field of view, a Windows OS USB insert sound is played, while an eject sound is played when the hand is out of the field of view. These sound cues help the user know when they can initiate interactions.

5 STUDY 2: COMPARING GUIDANCE SYSTEM AND PINPOINTING SYSTEM

The second study aimed to compare the effectiveness of the guidance and pinpointing systems in learning complex structures, specifically the ISS and Falcon 9. In addition to subjective ratings and feedback, we designed questions about the two models with input from the three experts who participated in Study 1. The questions evaluated objective learning outcomes in terms of textual and spatial understanding. The textual questions contained two parts, three questions examining textual information about composites and three questions examining basic elements. The spatial questions also had two parts, three area-related questions, and three location-related questions. Presented in Appendix A, these questions were prepared to test the following hypotheses:

- Hypothesis 1: Users would answer more textual questions correctly after using the guidance system because it systematically introduces information about both composites and basic elements. Conversely, the pinpointing system requires users to identify basic elements and then integrate the information to build an overall picture.
- Hypothesis 2: Users would answer more spatial questions correctly after using the guidance system because it supports more coherent spatial examination with navigation and confirmation functions. Conversely, the pinpointing system users must rely on discrete pinpointing to build spatial knowledge.

[4] https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html

Table 2: Participants' demographic information and their confidence in learning the topic.

ID	Age	Blind since	Confidence in learning the topic (1: Not confident at all 7: Highly confident)
P1	54	0	1
P2	20	0	1
P3	66	50	4
P4	36	14	1
P5	71	60	1
P6	50	37	1
P7	48	30	3
P8	30	10	5

Information in the guidance system reflects the complexity level for public learning found in 3.5.2. To compensate for the lack of information due to the pinpointing system's inability to present hierarchical layers, we included additional basic elements and ensured that all answers could be found in both systems.

5.1 Participants

We recruited eight blind participants (female = 4) who were unfamiliar with space technology such as the ISS, spacecraft, and rockets, and had never interacted with such models before. Their ages ranged from 20 to 71 years (mean = 46.8, SD = 17.5), and most of them rated themselves as having low confidence in learning the topic (median = 1), as listed in Table 2. Participants were recruited through an e-newsletter for individuals with visual impairments and were compensated for their time with $75 plus travel expenses. Seven out of eight participants had prior experience with 3D models, with participant P6 being the exception, and all of them stated that an explanation was necessary when examining the 3D models.

5.2 Procedure

5.2.1 Pre-Study Interview. Before introducing the system, we conducted an approximately 8-minute pre-study interview to inquire about participants' experience with 3D models, the need for explanations when examining the models, and their confidence in learning about space technologies, rated on a scale from 1 (not confident at all) to 7 (highly confident).

5.2.2 Training. The study began with a training session to familiarize participants with the systems. It involved presenting a 3D floor plan and conducting the following steps: (1) introducing the gestures and sounds used in the system, (2) introducing the pinpointing system and its stop button, while encouraging the participant to use the gesture to trigger audio labels within 5 minutes, and (3) introducing the guidance system and its next/previous buttons, and encouraging the participant to use the guidance system for about 5 minutes to familiarize themselves with the system. The entire training process took around 10 minutes.

5.2.3 Main Study. The main study used the ISS and Falcon 9 models, each with approximately 10 minutes of audio information for both the guidance and pinpointing systems. Participants started by

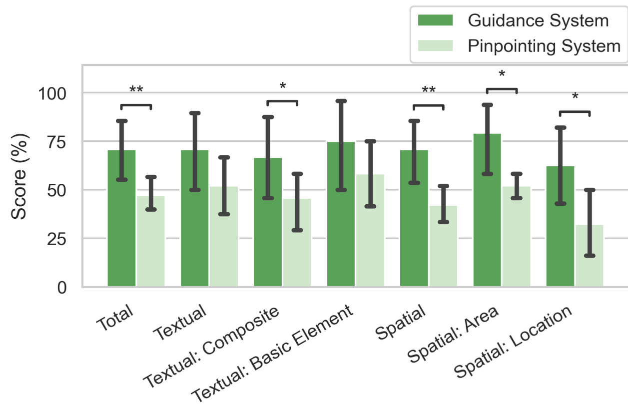


Figure 8: Mean of correct answer percentages to the questions depending on the system (and * indicate the 0.01 and 0.05 levels of one-tailed paired t-test p-value significance, respectively).**

exploring a model using the guidance system, answering model-related questions, and then exploring the same model again using the pinpointing system. Next, the other model was explored in reverse order, first using the pinpointing system and then the guidance system, with questions answered between these trials. Since the pinpointing system had no set end point, participants notified the experimenters when they had finished exploring. The order in which the models were presented was counterbalanced, with four participants starting with the ISS and four starting with Falcon 9. Before exploration, both systems provided the same model overview. The question-answering task took approximately 5 minutes.

5.2.4 Post-Study Interview. The interview consisted of four sections: (1) rating the level of independence and enjoyment for each system (Q1–Q2 in Table 3); (2) choosing the preferred way to use the systems in a real-world context and explaining why; (3) rating the usability of each specific function in terms of A. understandability and B. usefulness (Q3–Q7 in Table 4); and (4) free responses about the strengths and limitations of the systems, and other potential applications of the guidance system. This session took 20 minutes, and the entire study took approximately 1.5 hours.

5.3 Results

5.3.1 Time and Learning Outcome. The mean time using the guidance system was 17.75 minutes (SD = 9.13), while it was 6.88 minutes (SD = 2.53) with the pinpointing system. All participants spent more time with the guidance system, despite the length of information being the same for both systems. We noticed most people using the pinpointing system finished without discovering all elements. In contrast, with the guidance system, participants spent more time exploring spatially through navigation and confirmation.

The results of the questions are shown in Fig. 8. The mean of the total correct answer rate was 70.8% (SD=24.7%) with the guidance system and 47.1% (SD=12.8%) with the pinpointing system. Participants were able to answer significantly more questions correctly using the guidance system compared to the pinpointing system (t(7)=3.22, p<.01 (one-tailed)). We also found one participant (P3) answered all questions correctly after using the guidance system.

For all six sub-categories (textual questions, composite-related and basic element-related textual questions, spatial questions, area-related and location-related spatial questions), the mean correct answer rates were higher using the guidance system compared to the pinpointing system. We also found the higher correct answer rates significant for composite-related textual questions (t(7)=1.93, p<.05 (one-tailed)), all spatial questions (t(7)=3.20, p<.01 (one-tailed), area-related questions (t(7)=2.73, p<.05 (one-tailed)), and location-related questions (t(7)=2.28, p<.05 (one-tailed)).

5.3.2 Independence and Enjoyment. The ratings of independence and enjoyment (Q1–Q2) are summarized in Fig. 9, and the details are shown in Table 3. All participants agreed that the guidance system allowed them to learn the model from overviews to details independently (Q1.A, median = 6), while they were neutral toward the independence provided by the pinpointing system (Q1.B, median = 4). Most participants rated higher independence with the guidance system than with the pinpointing system, except for P5, who strongly agreed that both systems supported independent learning (Table 3). Furthermore, participants (P1, P2, P4, P6, and P8) provided comments indicating that the guidance system helped them to construct an overall understanding of the complex model. Conversely, they (P1, P3, P4, P6, P7) commented that they might miss information with the pinpointing system. One participant (P8) also noted that compared to the pinpointing system, the guidance system was especially helpful for hierarchical information.

> A1: *"The models are complicated, and I have never seen them before. Thus, it is impossible to touch everything (with the pinpointing system). In particular, the cylinder-shaped rocket seems to be simple, but its parts are small and intricate, which makes them difficult to find (using the pinpointing system). The guidance system guides me without omission."* P7
>
> A2: *"The ISS is complex. Using the pinpointing system, it's hard to tell whether something is a single element or part of a bigger structure."* P8

Four participants (P4, P6, P7, P8) praised the independence given by the guidance system, as they appreciated being able to control the pace of learning.

> A3: *"Usually, I need to be considerate not to bother the guide person too much. I am so happy to be able to touch to learn it alone at my own pace."* P6
>
> A4: *"I usually need to ask for explanations, but this system is user-friendly and allows me to examine the models closely without hesitation. I would be delighted if there were more available. Contents are usually simplified for the visually impaired; therefore, I am impressed that I can learn something really complex at my own pace with this system."* P8

Regarding enjoyment, most participants strongly agreed that the guidance system was enjoyable (Q2.A, median = 7), but they also agreed the pinpointing system was enjoyable (Q2.B, median = 6). Five participants reported higher enjoyment with the guidance system than with the pinpointing system, while others strongly agreed that both systems were enjoyable. Two participants (P6, P7) noted that the pinpointing system was less enjoyable than the

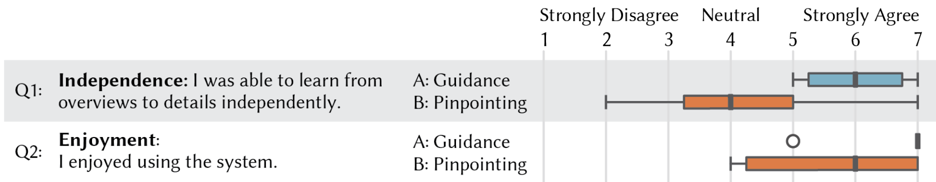


Figure 9: Questionnaire results of the independence (Q1) and enjoyment (Q2) for A: Guidance system, and B: Pinpointing system on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

guidance system because they needed to search for elements with audio labels proactively.

In contrast, one participant (P4) found both systems less enjoyable due to the complexity of the model and information presented.

> A5: *"The models are complex, and the information is hard to understand. I have tried hard to figure out the shape. Thus my mind does not allow me to process the (extended) information at the same time. It would be better if I were allowed to spend more time, one or two hours, to touch and go over the guidance several times."* P4

5.3.3 User Preferences. Participants were given five options to choose their preferred way to use the systems: (1) Guidance only, (2) Pinpointing only, (3) go through Guidance first, then Pinpointing, (4) explore with Pinpointing first, then Guidance, (5) Other. All participants chose (3) go through Guidance first, then Pinpointing. There was strong agreement among all participants that they preferred to go through the guidance to learn the overall image, then use pinpointing freely to confirm the knowledge.

> A6: *"It is easier for me to understand if I listen to the entire thing, and then review parts on an individual basis."* P3
> A7: *"I will omit things using the pinpointing system. But using the guidance I can comprehend everything from one end to the other. For review proposes, I can point to the part that I don't remember to confirm it. It would be great if I can switch between them."* P6

5.3.4 System Usability. The results of usability-related ratings (Q3.A–Q7.B) are summarized in Fig. 10 and the details are shown in Table 4. The participants agreed that the guidance system was easy to use (Q3.A, median = 6.5) and useful for understanding the model (Q3.B, median = 6). Correspondingly, they somewhat agreed the pinpointing system was easy to use (Q4.A, median = 5) and agreed it was useful (Q4.B, median = 6). As shown in Table 4, five participants (P1, P2, P3, P4, P7) rated higher easiness with the guidance system, while others rated both as easy to use (score = 6 or 7). Two participants (P2, P7) complained that they must use pointing gestures more frequently in the pinpointing system, which was difficult. Two other participants (P3, P4) pointed out that the pinpointing system was not as easy as the guidance system because it did not give timely confirmation.

> A8: *"Sometimes I pointed to a place, I expected it to speak but it did not. I didn't know whether that was due to no information or my pointing not being recognized."* P3

As for usefulness, participants' opinions were similar to their preferences. Two participants (P3, P8) commented that each of the two systems has its own usefulness.

> A9: *"Of course the guidance system is useful, but the pinpointing system is also useful because I can review the parts that caught my interest. Besides that, the guidance system is for those who are serious and want to take their time (in learning it), and the pinpointing system is suitable for those who want to do it casually."* P8

Participants somewhat agreed that the pointing gesture adopted in both systems was easy to use (Q5.A median = 5) and useful for understanding the model (Q5.B, median = 6). Three participants (P2, P5, P6) commented that the fingertip was not accurate enough, affecting their easiness rating.

> A10: *"At some locations, it was difficult to perform a nice pointing gesture that could be recognized by the camera. I'd like something like a touch pen that can accurately point to fine details."* P2

We also observed that while four participants (P3, P6, P7, P8) skillfully switched between normal tactile examination and the pointing gesture, the others (P1, P2, P4, P5) tended to keep using the pointing gesture all the time. One participant (P4) criticized the pointing gesture as tiring and difficult. However, two participants who switched between gesture and regular exploration (P7, P8) commented the opposite.

> A11: *"At first, I wasn't used to (switching the gestures). But I gradually got used to switching between gestures and spreading my fingers (to avoid occlusion), and they weren't difficult for me anymore."* P7
> A12: *"The nice part was that when I explored with both hands, the system did not say anything, so I could concentrate on touching. I could point at the right time to listen to the audio."* P8

As for functions only available in the guidance system, they agreed the navigation was easy to use (Q6.A, median = 6) and useful for understanding the model (Q6.B, median = 6.5). Two participants (P2, P4) pointed out that the navigation was not precise when the finger approached a finely detailed place.

> A13: *"It told me to go down, then right, then up again. I got confused because I didn't know which direction to go."* P4

They also agreed that the confirmation sound was somewhat easy to use (Q7.A, median = 5.5) and useful (Q7.B, median = 6.5). They gave feedback that accuracy was a factor affecting the easiness. Two participants (P2, P4) pointed out that the sound indication was not precise at some locations. Two other participants (P5, P6)

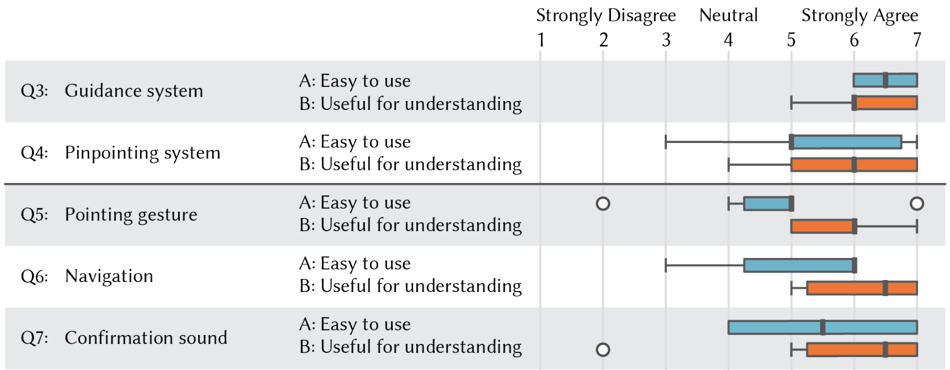


Figure 10: Questionnaire results of the usability of the two systems (Q3–Q4), pointing gesture that was used in both systems (Q5), and functions in the guidance system (Q6-Q7) on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). A: [It is] easy to use, and B: [It is] useful for understanding [the model structure.]

commented that the sound was nearly unnoticeable, which affected the scores. As for usefulness, several participants (P3, P6, P7, P8) expressed the feeling that it helped to grasp the area that was explained.

> A14: *"The model is in three dimensions. But sometimes the (occluded) sides and bottoms did not activate the sound."* P4
> A15: *"It would be better if the different sounds were easier to distinguish. However, the confirmation sound was helpful to know whether I (entered or) left the correct area."* P6

5.3.5 Suggestions. Two participants (P2, P7) reported that the camera occlusion limited their ability to touch certain parts of the model, particularly the bottom and front faces, which were obscured by the model itself. Additionally, two other participants (P1, P8) noted that the reliability was decreased when the functions were not accurate.

> A16: *"I didn't know how much to believe when (it) was off at some part."* P1
> A17: *"I want more explanations of the shapes. Then I could trust that I'm touching the correct location."* P8

One participant hoped to increase the modes to learn different aspects of the model.

> A18: *"After understanding the overall picture, I want to learn more characteristics, such as color, material, temperature, and warmth. It would be helpful to have more modes."* P5

Participants also provided insights into the potential applications of the system. The answers are categorized as follows: artifacts in museums that need explanations (P2, P3, P4, P5, P8), the structure of architecture and maps (P2, P3, P4), objects that are too large or far away to touch (P4, P6, P7), and complex interfaces in their daily life, such as the computer window layout (P1) and the user interface of home appliances (P8).

6 DISCUSSION

6.1 Effectiveness of the Guidance System

Answering the research question of *what kind of I3Ms can assist blind people in learning complex structures independently,* we found that the guidance system effectively assisted the user in the following three aspects: (1) better objective learning outcome textually and spatially; (2) higher independence in obtaining the overall picture and the details; (3) ease and usefulness of navigating, locating, and confirming an area, which helped the independent learning.

As for the objective learning outcomes, we hypothesized that the guidance system would support better textual and spatial understanding and memorization. The results of question answering mostly supported this because the participants using the guidance system had higher mean correct answers in all categories of questions. The trends are significant for spatial questions, including area-related and location-related ones, and composite-related textual questions. Currently, the guidance system is not significantly more robust in supporting basic element-related textual questions. We assume that since the pinpointing system was single-layered, it might not be weaker in providing such fundamental information. Still, the significant differences need to be confirmed with more users. Meanwhile, qualitative findings support better learning outcomes with the guidance system, as users commented it promoted identifying an element (A1, A2), grasping the shape and area (A15), and comprehending an overall image (A6, A7). The longer duration of usage and the observation that users studied the model thoroughly using the guidance system, compared to missing information using the pinpointing system, also reflect higher engagement, effectiveness, and potential learning gains.

Being able to obtain knowledge using the guidance system also leads to high independence, as participants rated the guidance system as independent and were neural about the independence provided by the pinpointing system. Participants found that the guidance system better helped them to build a comprehensive understanding from hierarchies (A2) to details (A1). They also commented that such independence was crucial for them, as they could control the learning pace by themselves (A3, A4). Both Study 1 observations (in Section 3.5.4) and the above comments reveal that independence in learning a complex structure did not merely mean acquiring information by oneself. It also involved obtaining an overall picture if the user did not have prior knowledge, controlling the pace of learning, and then being able to acquire information freely.

The guidance system served as a primer for learning the complex structure.

All functions in the guidance system were useful in supporting independent and effective learning of a complex structure. The ordered guide supported the development of the big picture and hierarchical knowledge. The navigation helped the user to locate the target element and build spatial knowledge. The confirmation double-confirmed the spatial location and helped grasp the entire area during regular touch examinations. Most participants agreed these functions were easy to use, and those who criticized them revealed that the accuracy was the biggest reason affecting the easiness (A13, A14). Switching between the regular exploration style and the pointing gesture might need some practice (A11), but it was preferred over pointing all the time (A12).

6.2 Integrating Guidance and Pinpointing Systems Sequentially

One interesting finding is that, although both systems provide similar information, all of the participants preferred to use the system in the same order: guidance first, then pinpointing. None chose to use one system alone or in the reverse order. Several reasons could account for this. First, the model was complex and unfamiliar. Participants needed to be primed with a structured overview before freely inquiring about information. Moreover, a review was needed for such complex information, since memory and comprehension played an essential role in learning. Ideally, after using the guide system, the user could answer all of the questions correctly if they understood and memorized all of the information. However, only one participant (P3) achieved this. Several participants specified that a pipeline of being primed by the guide and then reviewing elements freely was the best way to learn such complex structure (A6, A7, A9). Both systems were rated enjoyable, suggesting they can be integrated to provide an enjoyable learning experience combining structured learning and free exploration.

Participants also suggested switching between modes (A7) and repeating the guidance (A5) by themselves to obtain maximum autonomy in learning. One participant suggested including more characteristics after using the guidance (A18). This confirmed the finding of Study 1 that participants asked follow-up questions after hearing the explanation. Thus, integrating supplementary modes, which include but are not limited to the pinpointing system, can be helpful to give the user freedom to acquire more knowledge. Since Study 1 also revealed that contextual information was a part of the knowledge about a subject, a conversational agent (e.g., [45]) could also be helpful after the guidance system laid the main groundwork.

6.3 Limitations and Future Work

The biggest limitation of the systems is the accuracy of hand tracking using the optical and depth camera-based approach, which affected pointing, navigation, and confirmation functions. When one hand is occluded by other objects or the other hand, the interaction becomes unstable. We used USB insertion and ejection sounds to indicate the trackable and untrackable states; however, these sounds might be inadequate and difficult to distinguish (A15). From user experience, we found occlusion issues included untrackable areas (A14) and inaccuracies for small areas (A13) and fine details (A10). The instability also decreased the system's reliability (A8, A16, A17). Thus, the first technical challenge is to either increase the accuracy or clearly indicate when the accuracy drops. Higher accuracy could be achieved by improving the hardware setup to minimize occlusion or obtaining more accurate 3D hand data from other sensors. To indicate the drop in accuracy, we could add another classifier to the hand detection to check whether the hand is occluded.

Since participants expressed interest in having more models and even more modes available through the system (Section 5.3.5), further generalization should be investigated. Potentially, the base system could be further improved to detect and track customized objects, allowing interactions with moving models. Advanced controls like gestures on both hands should also be explored. Even though our approach has only been implemented in a limited number of space engineering-related models and tested with individuals who did not necessarily have a particular interest in space technology, it shows potential for learning other complex structures both individually and in groups. Nowadays, group learning involves a lecture followed by sighted users helping blind users to review the knowledge [58]. We should investigate the effectiveness of introducing our system to guide a group of learners. Furthermore, another avenue for future research is to explore how experts and teachers can create customized I3Ms. Currently, we created fixed textual content with the help of the experts. However, some participants expressed the need for customized content, such as descriptions of shapes (A17) and more characteristics (A18). Nowadays, as large language models (LLMs) become skillful at explaining things [39], we can examine how this artificial intelligence can assist in creating various guide content. In our next step, we need to query not only the visually impaired but also experts and teachers to investigate more scenarios (e.g., classrooms), other system functions (e.g., bimanual controls and conversational agents), and a simplified pipeline of model preparation and interactive content creation.

7 CONCLUSION

In this paper, we introduced TouchPilot, a guidance system that makes complex and unfamiliar 3D structures accessible for blind people. We first conducted a study to investigate natural learner-explainer interaction styles in learning complex structures. Based on the observations and feedback, we designed TouchPilot to guide the user through exploring hierarchical elements, navigating the elements using directional cues, and confirming the area of the target element using sound. A follow-up study comparing the guidance system with the existing pinpointing-based interactive 3D printed model system reveals that the guidance system led to improved learning outcomes regarding hierarchical textual and spatial information as well as increased independence. The study also probed real-world usage, finding that participants preferred to use TouchPilot initially to establish a foundation and then employing the pinpointing system or other modes to review and deepen their understanding freely. We hope our findings will promote further explorations of how to make complex and educational information accessible to the visually impaired.

ACKNOWLEDGMENTS

We thank all participants who took part in our user studies. We also thank Hitoshi Matsuoka, Chief Science Communication Specialist, as well as Kodai Sano, Natsumi Nakano, Ruriko Nagashima, and Bunsuke Kawasaki, the Science Communicators at Miraikan – The National Museum of Emerging Science and Innovation, for their contributions to the system design and related studies.

REFERENCES

[1] Frances K Aldrich and Linda Sheppard. 2001. Tactile graphics in school education: perspectives from pupils. *British Journal of Visual Impairment* 19, 2 (2001), 69–73.

[2] Kimberly Arcand, April Jubett, Megan Watzke, Sara Price, Kelly Williamson, and Peter Edmonds. 2019. Touching the stars: improving NASA 3D printed data sets with blind and visually impaired audiences. *Journal of Science Communication* 18, 04 (July 2019), A01. https://doi.org/10.22323/2.18040201

[3] Felix W. Baumann and Dieter Roller. 2018. Thingiverse: review and analysis of available files. *International Journal of Rapid Manufacturing* 7, 1 (2018), 83. https://doi.org/10.1504/ijrapidm.2018.089731

[4] Anke M. Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2014. Interactivity Improves Usability of Geographic Maps for Visually Impaired People. *Human–Computer Interaction* 30, 2 (Nov. 2014), 156–194. https://doi.org/10.1080/07370024.2014.924412

[5] Emeline Brulé and Gilles Bailly. 2021. ”Beyond 3D Printers”: Understanding Long-Term Digital Fabrication Practices for the Education of Visually Impaired or Blind Youth. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI ’21)*. Association for Computing Machinery, New York, NY, USA, Article 44, 15 pages. https://doi.org/10.1145/3411764.3445403

[6] Erin Buehler, Stacy Branham, Abdullah Ali, Jeremy J. Chang, Megan Kelly Hofmann, Amy Hurst, and Shaun K. Kane. 2015. Sharing is Caring: Assistive Technology Designs on Thingiverse. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI ’15)*. Association for Computing Machinery, New York, NY, USA, 525–534. https://doi.org/10.1145/2702123.2702525

[7] Erin Buehler, Niara Comrie, Megan Hofmann, Samantha McDonald, and Amy Hurst. 2016. Investigating the Implications of 3D Printing in Special Education. *ACM Trans. Access. Comput.* 8, 3, Article 11 (mar 2016), 28 pages. https://doi.org/10.1145/2870640

[8] Erin Buehler, Niara Comrie, Megan Hofmann, Samantha McDonald, and Amy Hurst. 2016. Investigating the Implications of 3D Printing in Special Education. *ACM Trans. Access. Comput.* 8, 3, Article 11 (mar 2016), 28 pages. https://doi.org/10.1145/2870640

[9] Francesco Buonamici, Rocco Furferi, Lapo Governi, and Yary Volpe. 2015. Making Blind People Autonomous in the Exploration of Tactile Models: A Feasibility Study. In *Universal Access in Human-Computer Interaction. Access to Interaction*, Margherita Antona and Constantine Stephanidis (Eds.). Springer International Publishing, Cham, 82–93.

[10] James M Coughlan, Huiying Shen, and Brandon Biggs. 2020. Towards accessible audio labeling of 3D objects. In *Journal on technology and persons with disabilities:... Annual International Technology and Persons with Disabilities Conference*, Vol. 8. NIH Public Access, 210.

[11] F. D’Agnano, C. Balletti, F. Guerra, and P. Vernier. 2015. Tooteko: A case study of augmented reality for an accessible cultural heritage. Digitization, 3D printing and sensors for an audio-tactile experience. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W4 (2015), 207–213. https://doi.org/10.5194/isprsarchives-XL-5-W4-207-2015

[12] Yvonne Eriksson. 1999. How to make tactile pictures understandable to the blind reader. In *Proceedings of the 65th IFLA Council and General Conference.*

[13] Giovanni Fusco and Valerie S. Morash. 2015. The Tactile Graphics Helper: Providing Audio Clarification for Tactile Graphics Using Machine Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers Accessibility* (Lisbon, Portugal) *(ASSETS ’15)*. Association for Computing Machinery, New York, NY, USA, 97–106. https://doi.org/10.1145/2700648.2809868

[14] Stéphanie Giraud, Anke M Brock, Marc J-M Macé, and Christophe Jouffrais. 2017. Map learning with a 3D printed interactive small-scale model: Improvement of space and text memorization in visually impaired students. *Frontiers in psychology* 8 (2017), 930.

[15] Stéphanie Giraud, Philippe Truillet, Véronique Gaildrat, and Christophe Jouffrais. 2017. “DIY” Prototyping of Teaching Materials for Visually Impaired Children: Usage and Satisfaction of Professionals. In *Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I 11*. Springer, 515–524.

[16] Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning Non-Visual Graphical Information Using a Touch-Based Vibro-Audio Interface. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) *(ASSETS ’12)*. Association for Computing Machinery, New York, NY, USA, 103–110. https://doi.org/10.1145/2384916.2384935

[17] Timo Götzelmann. 2016. LucentMaps: 3D Printed Audiovisual Tactile Maps for Blind and Visually Impaired People. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS ’16)*. Association for Computing Machinery, New York, NY, USA, 81–90. https://doi.org/10.1145/2982142.2982163

[18] William Grussenmeyer, Jesel Garcia, and Fang Jiang. 2016. Feasibility of Using Haptic Directions through Maps with a Tablet and Smart Watch for People Who Are Blind and Visually Impaired. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI ’16)*. Association for Computing Machinery, New York, NY, USA, 83–89. https://doi.org/10.1145/2935334.2935367

[19] Jaume Gual, Marina Puyuelo, Joaquim Lloveras, et al. 2011. Universal design and visual impairment: Tactile products for heritage access. In *Proceedings of the 18th International Conference on Engineering Design (ICED 11)*. Lyngby/Copenhagen, Denmark, 155–164.

[20] Jaume Gual, Marina Puyuelo, Joaquim Lloverás, and Lola Merino. 2012. Visual impairment and urban orientation. Pilot study with tactile maps produced through 3D printing. *Psyecology* 3, 2 (2012), 239–250.

[21] Anhong Guo, Xiang ’Anthony’ Chen, Haoran Qi, Samuel White, Suman Ghosh, Chieko Asakawa, and Jeffrey P. Bigham. 2016. VizLens: A Robust and Interactive Screen Reader for Interfaces in the Real World. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST ’16)*. Association for Computing Machinery, New York, NY, USA, 651–664. https://doi.org/10.1145/2984511.2984518

[22] Simon J Hayhoe. 2014. An enquiry into passive and active exclusion from unreachable artworks in the museum: Two case studies of final-year students at California School for the Blind studying artworks through galleries and on the web. *British Journal of Visual Impairment* 32, 1 (2014), 44–58.

[23] Leona Holloway, Kim Marriott, and Matthew Butler. 2018. Accessible Maps for the Blind: Comparing 3D Printed Models with Tactile Graphics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173772

[24] Leona Holloway, Kim Marriott, Matthew Butler, and Samuel Reinders. 2019. 3D Printed Maps and Icons for Inclusion: Testing in the Wild by People Who Are Blind or Have Low Vision. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS ’19)*. Association for Computing Machinery, New York, NY, USA, 183–195. https://doi.org/10.1145/3308561.3353790

[25] Amy Hurst and Jasmine Tobias. 2011. Empowering Individuals with Do-It-Yourself Assistive Technology. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) *(ASSETS ’11)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/2049536.2049541

[26] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (apr 1991), 152–169. https://doi.org/10.1145/123078.128728

[27] M. Gail Jones, James Minogue, Tom Oppewal, Michelle P. Cook, and Bethany Broadwell. 2006. Visualizing Without Vision at the Microscale: Students With Visual Impairments Explore Cells With Touch. *Journal of Science Education and Technology* 15, 5-6 (Nov. 2006), 345–351. https://doi.org/10.1007/s10956-006-9022-6

[28] Shaun K. Kane, Brian Frey, and Jacob O. Wobbrock. 2013. Access Lens: A Gesture-Based Screen Reader for Real-World Documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI ’13)*. Association for Computing Machinery, New York, NY, USA, 347–350. https://doi.org/10.1145/2470654.2470704

[29] Michael A Kolitsky. 2014. 3D printed tactile learning objects: Proof of concept. *Journal of Blindness Innovation and Research* 4, 1 (2014), 4–51.

[30] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng, and Anhong Guo. 2022. Image-Explorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI ’22)*. Association for Computing Machinery, New York, NY, USA, Article 462, 15 pages. https://doi.org/10.1145/3491102.3501966

[31] Barbara Leporini, Valentina Rossetti, Francesco Furfari, Susanna Pelagatti, and Andrea Quarta. 2020. Design Guidelines for an Interactive 3D Model as a Supporting Tool for Exploring a Cultural Site by Visually Impaired and Sighted People. *ACM Trans. Access. Comput.* 13, 3, Article 9 (aug 2020), 39 pages. https://doi.org/10.1145/3399679

[32] Jerzy Montusiewicz, Marcin Barszcz, and Sylwester Korga. 2022. Preparation of 3D Models of Cultural Heritage Objects to Be Recognised by Touch by the Blind—Case Studies. *Applied Sciences* 12, 23 (Nov. 2022), 11910. https://doi.org/10.3390/app122311910

[33] Jerzy Montusiewicz, Marek Miłosz, and Jacek Kęsik. 2018. Technical aspects of museum exposition for visually impaired preparation using modern 3D technologies. In *2018 IEEE Global Engineering Education Conference (EDUCON)*. 768–773. https://doi.org/10.1109/EDUCON.2018.8363308

[34] The Braille Authority of North America. 2010. Guidelines and Standards for Tactile Graphics. http://www.brailleauthority.org/tg/web-manual/index.html

[35] National Federation of the Blind. 2009. The Braille Literacy Crisis in America: Facing the Truth, Reversing the Trend, Empowering the Blind. https://nfb.org/images/nfb/documents/pdf/braille_literacy_report_web.pdf

[36] Mahika Phutane, Julie Wright, Brenda Veronica Castro, Lei Shi, Simone R. Stern, Holly M. Lawson, and Shiri Azenkot. 2022. Tactile Materials in Practice: Understanding the Experiences of Teachers of the Visually Impaired. *ACM Trans. Access. Comput.* 15, 3, Article 17 (jul 2022), 34 pages. https://doi.org/10.1145/3508364

[37] Thomas Poon and Ronit Ovadia. 2008. Using Tactile Learning Aids for Students with Visual Impairments in a First-Semester Organic Chemistry Course. *Journal of Chemical Education* 85, 2 (Feb. 2008), 240. https://doi.org/10.1021/ed085p240

[38] Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. 2023. Take My Hand: Automated Hand-Based Spatial Guidance for the Visually Impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 544, 16 pages. https://doi.org/10.1145/3544548.3581415

[39] Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Explain Yourself! Leveraging Language Models for Commonsense Reasoning. *CoRR* abs/1906.02361 (2019). arXiv:1906.02361 http://arxiv.org/abs/1906.02361

[40] Gaspar Ramôa, Vincent Schmidt, and Peter König. 2022. Developing Dynamic Audio Navigation UIs to Pinpoint Elements in Tactile Graphics. *Multimodal Technologies and Interaction* 6, 12 (2022). https://doi.org/10.3390/mti6120113

[41] Gaspar Ramôa, Vincent Schmidt, and Peter König. 2022. Developing Dynamic Audio Navigation UIs to Pinpoint Elements in Tactile Graphics. *Multimodal Technologies and Interaction* 6, 12 (2022). https://doi.org/10.3390/mti6120113

[42] Andreas Reichinger, Helena Garcia Carrizosa, Joanna Wood, Svenja Schröder, Christian Löw, Laura Rosalia Luidolt, Maria Schimkowitsch, Anton Fuhrmann, Stefan Maierhofer, and Werner Purgathofer. 2018. Pictures in Your Mind: Using Interactive Gesture-Controlled Reliefs to Explore Art. *ACM Trans. Access. Comput.* 11, 1, Article 2 (mar 2018), 39 pages. https://doi.org/10.1145/3155286

[43] Andreas Reichinger, Anton Fuhrmann, Stefan Maierhofer, and Werner Purgathofer. 2016. Gesture-Based Interactive Audio Guide on Tactile Reliefs. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) *(ASSETS '16)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/2982142.2982176

[44] Andreas Reichinger, Svenja Schröder, Christian Löw, Sam Sportun, Peter Reichl, and Werner Purgathofer. 2016. Spaghetti, Sink and Sarcophagus: Design Explorations of Tactile Artworks for Visually Impaired People. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 82, 6 pages. https://doi.org/10.1145/2971485.2996471

[45] Samuel Reinders, Matthew Butler, and Kim Marriott. 2020. "Hey Model!" – Natural User Interactions and Agency in Accessible Interactive 3D Models. 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, 1–13. https://doi.org/10.1145/3313831.3376145

[46] Audrey C. Rule. 2011. Tactile Earth and Space Science Materials for Students with Visual Impairments: Contours, Craters, Asteroids, and Features of Mars. *Journal of Geoscience Education* 59, 4 (Nov. 2011), 205–218. https://doi.org/10.5408/1.3651404

[47] Huiying Shen, Owen Edwards, Joshua Miele, and James M. Coughlan. 2013. CamIO: A 3D Computer Vision System Enabling Audio/Haptic Interaction with Physical Objects by Blind Users. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) *(ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, Article 41, 2 pages. https://doi.org/10.1145/2513383.2513423

[48] Lei Shi, Holly Lawson, Zhuohao Zhang, and Shiri Azenkot. 2019. Designing Interactive 3D Printed Models with Teachers of the Visually Impaired. 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, 1–14. https://doi.org/10.1145/3290605.3300427

[49] Lei Shi, Idan Zelzer, Catherine Feng, and Shiri Azenkot. 2016. Tickers and Talker: An Accessible Labeling Toolkit for 3D Printed Models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4896–4907. https://doi.org/10.1145/2858036.2858507

[50] Lei Shi, Yuhang Zhao, and Shiri Azenkot. 2017. Designing Interactions for 3D Printed Models with Blind People. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 200–209. https://doi.org/10.1145/3132525.3132549

[51] Lei Shi, Yuhang Zhao, and Shiri Azenkot. 2017. Markit and Talkit: A Low-Barrier Toolkit to Augment 3D Printed Models with Audio Annotations. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 493–506. https://doi.org/10.1145/3126594.3126650

[52] Lei Shi, Yuhang Zhao, Ricardo Gonzalez Penuela, Elizabeth Kupferstein, and Shiri Azenkot. 2020. Molder: An Accessible Design Tool for Tactile Maps. 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. https://doi.org/10.1145/3313831.3376431

[53] Ryo Suzuki, Abigale Stangl, Mark D. Gross, and Tom Yeh. 2017. FluxMarker: Enhancing Tactile Graphics with Dynamic Tactile Markers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 190–199. https://doi.org/10.1145/3132525.3132548

[54] Stella Sylaiou and Christos Fidas. 2022. Supporting People with Visual Impairments in Cultural Heritage: Survey and Future Research Directions. *International Journal of Human–Computer Interaction* (July 2022), 1–16. https://doi.org/10.1080/10447318.2022.2098930

[55] Carlos Tejada, Osamu Fujimoto, Zhiyuan Li, and Daniel Ashbrook. 2018. Blowhole: Blowing-Activated Tags for Interactive 3D-Printed Models. *Proceedings of Graphics Interface 2018* Toronto (2018), 131–137. https://doi.org/10.20380/GI2018.18

[56] Yoshinori Teshima, Atsushi Matsuoka, Mamoru Fujiyoshi, Yuji Ikegami, Takeshi Kaneko, Susumu Oouchi, Yasunari Watanabe, and Kenji Yamazawa. 2010. Enlarged skeleton models of plankton for tactile teaching. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July14-16, 2010, Proceedings, Part II 12*. Springer, 523–526.

[57] Lauren Thevin and Anke M. Brock. 2018. Augmented Reality for People with Visual Impairments: Designing and Creating Audio-Tactile Content from Existing Objects. In *Computers Helping People with Special Needs*, Klaus Miesenberger and Georgios Kouroupetroglou (Eds.). Springer International Publishing, Cham, 193–200.

[58] Kumiko Usuda-Sato, Hirotaka Nakayama, Hideaki Fujiwara, and Tomonori Usuda. 2019. Touch the Universe: Developing and Disseminating Tactile Telescope Models Created with a 3D Printer. *Communicating Astronomy with the Public Journal* 26, 4 (2019), 24–30.

[59] Roberto Vaz, Paula Odete Fernandes, and Ana Cecília Rocha Veiga. 2018. Designing an Interactive Exhibitor for Assisting Blind and Visually Impaired Visitors in Tactile Exploration of Original Museum Pieces. *Procedia Computer Science* 138 (2018), 561–570. https://doi.org/10.1016/j.procs.2018.10.076 CENTERIS 2018 - International Conference on ENTERprise Information Systems / ProjMAN 2018 - International Conference on Project MANagement / HCist 2018 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018.

[60] Andreas Voigt and Bob Martens. 2006. Development of 3D tactile models for the partially sighted to facilitate spatial orientation. In *24th eCAADe Conference Proceedings*. CUMINCAD.

[61] Xi Wang, Danny Crookes, Sue-Ann Harding, and David Johnston. 2022. Stories, journeys and smart maps: an approach to universal access. *Universal Access in the Information Society* 21, 2 (2022), 419–435.

[62] Xiyue Wang, Seita Kayukawa, Hironobu Takagi, and Chieko Asakawa. 2022. BentoMuseum: 3D and Layered Interactive Museum Map for Blind Visitors. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 14 pages. https://doi.org/10.1145/3517428.3544811

[63] Henry B. Wedler, Sarah R. Cohen, Rebecca L. Davis, Jason G. Harrison, Matthew R. Siebert, Dan Willenbring, Christian S. Hamann, Jared T. Shaw, and Dean J. Tantillo. 2012. Applied Computational Chemistry for the Blind and Visually Impaired. *Journal of Chemical Education* 89, 11 (Sept. 2012), 1400–1404. https://doi.org/10.1021/ed3000364

[64] David Wicks. 2017. The coding manual for qualitative researchers. *Qualitative research in organizations and management: an international journal* 12, 2 (2017), 169–170.

[65] Andrew D. Wilson. 2010. Using a Depth Camera as a Touch Sensor. In *ACM International Conference on Interactive Tabletops and Surfaces* (Saarbrücken, Germany) *(ITS '10)*. Association for Computing Machinery, New York, NY, USA, 69–72. https://doi.org/10.1145/1936652.1936665

[66] Paul F. Wilson, Susan Griffiths, Ellis Williams, M. Paul Smith, and Mark A. Williams. 2020. Designing 3-D Prints for Blind and Partially Sighted Audiences in Museums: Exploring the Needs of Those Living with Sight Loss. *Visitor Studies* 23, 2 (June 2020), 120–140. https://doi.org/10.1080/10645578.2020.1776562

[67] Paul F. Wilson, Janet Stott, Jason M. Warnett, Alex Attridge, M. Paul Smith, and Mark A. Williams. 2017. Evaluation of Touchable 3D-Printed Replicas in Museums. *Curator: The Museum Journal* 60, 4 (Oct. 2017), 445–465. https://doi.org/10.1111/cura.12244

[68] Jo Wonjin, Hee I Jang, Rachel Ananda Harianto, Ji Hyun So, Hyebin Lee, Heon Ju Lee, and Myoung-Woon Moon. 2016. Introduction of 3D printing technology in the classroom for visually impaired students. *Journal of Visual Impairment & Blindness* 110, 2 (2016), 115–121.

[69] Tessa Wright, Beth Harris, and Eric Sticken. 2010. A best-evidence synthesis of research on orientation and mobility involving tactile maps and models. *Journal of visual impairment & blindness* 104, 2 (2010), 95–106.
[70] Kim T Zebehazy and Adam P Wilton. 2014. Quality, importance, and instruction: The perspectives of teachers of students with visual impairments on graphics use by students. *Journal of visual impairment & blindness* 108, 1 (2014), 5–16.
[71] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).
[72] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. 2018. Open3D: A modern library for 3D data processing. *arXiv preprint arXiv:1801.09847* (2018).

A QUESTIONS TO EVALUATE THE LEARNING OUTCOME

The questions were translated from the participants' native language. The question type was indicated by angle brackets, and a score was assigned to each answer, enclosed in parentheses. Scores were normalized to percentages for analysis purposes. An announcement was made before answering the questions: "We prepared six multiple-choice questions and six model-pointing questions to test your knowledge. Please answer them based on what you learned from the model. 'I don't know' is also a choice, so you don't need to guess."

A.1 ISS

A.1.1 Part 1. Textual information.

(1) <Basic element> How many solar panels does the ISS have?
A. 4 (0) B. 8 (2) C. I don't know (0)
(2) <Basic element> Japan laboratory module Kibo and U.S. laboratory Destiny, which one is bigger?
A. Kibo (2) B. Destiny (0) C. I don't know (0)
(3) <Basic element> Is there a balloon-like module in ISS?
A. No (0) B. Yes (2) C. I don't know (0)
(4) <Composite> Which country does the first module launched to the ISS belong to?
A. U.S. (0) B. Russia (2) C. I don't know (0)
(5) <Composite> Which of the following is used as a laboratory and living space?
A. Truss (0) B. Pressurized module (2) C. I don't know (0)
(6) <Composite> What does the ISS airlock mean?
A. Entrance and exit for EVA activities (2) B. Device to control oxygen and carbon dioxide saturation (0) C. I don't know (0)

A.1.2 Part 2. Spatial information.

(1) <Area> Point me to the entire shape of the Truss.
A. Correct (2) B. Partially correct (1) C. Wrong (0)
(2) <Area>Point me to the entire shape of the Russian modules.
A. Correct (2) B. Partially correct (1) C. Wrong (0)
(3) <Area> Point me to the entire shape of the Japan modules.
A. Correct (2) B. Partially correct (1) C. Wrong (0)
(4) <Location> Point me where the spacecraft Crew Dragon can dock on this model.
A. Correct (2) B. Wrong (0)
(5) <Location> Point me where is the module that has seven windows.
A. Correct (2) B. Wrong (0)
(6) <Location> Point me to all places where extra-vehicular activities can be done. (3 places)
A. Three correct objects (3) B. Two correct objects (2) C. One correct object (1) D. Wrong (0)

A.2 Falcon 9

A.2.1 Part 1. Textual information.

(1) <Basic element> Falcon 9 is made by what kind of company?
A. The American space agency (0) B. An American private company (2) C. I don't know (0)
(2) <Basic element> Why is this rocket named Falcon 9?
A. Stage 1 has 9 engines (2) B. It is the ninth version of the Falcon rocket (0) C. I don't know (0)
(3) <Basic element> What name is the drone ship under the Falcon 9 model?
A. Of Course I Still Love You (2) B. Just Read The Instructions (0) C. I don't know (0)
(4) <Composite> How many stages does Falcon 9 have?
A. 2 (2) B. 3 (0) C. I don't know (0)
(5) <Composite> Which parts of Falcon 9 are recovered and reused?
A. Stage1 (2) B. All parts of Crew Dragon craft (capsule and trunk) (0) C. I don't know (0)
(6) <Composite> What can spacecraft Crew Dragon load?
A. more than 8 people (0) B. A large amount of cargo (2) C. I don't know (0)

A.2.2 Part 2. Spatial information.

(1) <Area> Point me to the entire area of the Crew Dragon.
A. Correct (2) B. Partially correct (1) C. Wrong (0)
(2) <Area> Point me from where to where the Fuel tank is.
A. Correct (2) B. Partially correct (1) C. Wrong (0)
(3) <Area> Stage 1 and Stage 2, which one is longer?
A. Correct (2) B. Wrong (0)
(4) <Location> Point me where the separation point of Stages 1 and Stages 2 is.
A. Correct (2) B. Wrong (0)
(5) <Location> Point me where the SpaceX logo is.
A. Correct (2) B. Wrong (0)
(6) <Location> Point me to all three objects stage 1 needs for the landing.
A. Three correct objects (3) B. Two correct objects (2) C. One correct object (1) D. Wrong (0)

B INDIVIDUAL RESPONSES TO THE QUESTIONNAIRE

Table 3: Individual responses and the median to the independence (Q1) and enjoyment (Q2) for A: Guidance system, and B: Pinpointing system on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

Question	System	P1	P2	P3	P4	P5	P6	P7	P8	Median
Q1. Independence: I was able to learn from overviews to details independently.	A. Guidance	6	6	6	5	7	5	6	7	6
	B. Pinpointing	4	4	4	2	7	3	5	5	4
Q2. Enjoyment: I enjoyed using the system.	A. Guidance	7	7	7	5	7	7	7	7	7
	B. Pinpointing	5	4	7	4	7	6	6	7	6

Table 4: Individual responses and the median to the usability of the two systems (Q3–Q4), the pointing gesture used in both systems (Q5), and functions in the guidance system (Q6-Q7) on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). A: [It is] easy to use, and B: [It is] useful for understanding [the model structure.]

Question		P1	P2	P3	P4	P5	P6	P7	P8	Median
Q3. Guidance system	A. Easy to use	7	6	6	6	7	6	7	7	6.5
	B. Useful for understanding	7	6	6	5	7	6	6	7	6
Q4. Pinpointing system	A. Easy to use	5	3	5	5	7	6	5	7	5
	B. Useful for understanding	5	5	6	4	7	7	6	7	6
Q5. Pointing gesture	A. Easy to use	5	4	5	2	5	5	5	7	5
	B. Useful for understanding	6	5	5	5	6	6	6	7	6
Q6. Navigation	A. Easy to use	6	4	6	3	6	6	6	5	6
	B. Useful for understanding	7	5	6	5	7	7	6	7	6.5
Q7. Confirmation sound	A. Easy to use	7	4	6	1	4	5	7	7	5.5
	B. Useful for understanding	7	5	6	2	6	7	7	7	6.5