# Don't Look Now: Audio/Haptic Guidance for 3D Scanning of Landmarks

Jessica Van Brummelen
Liv Piper Urwin
Oliver James Johnston
Mohamed Sayed
Gabriel Brostow
jvanbrummelen@nianticlabs.com
livurwin@nianticlabs.com
oliverjohnston@nianticlabs.com
mohameds@nianticlabs.com
gbrostow@nianticlabs.com
Niantic Inc.
London, UK

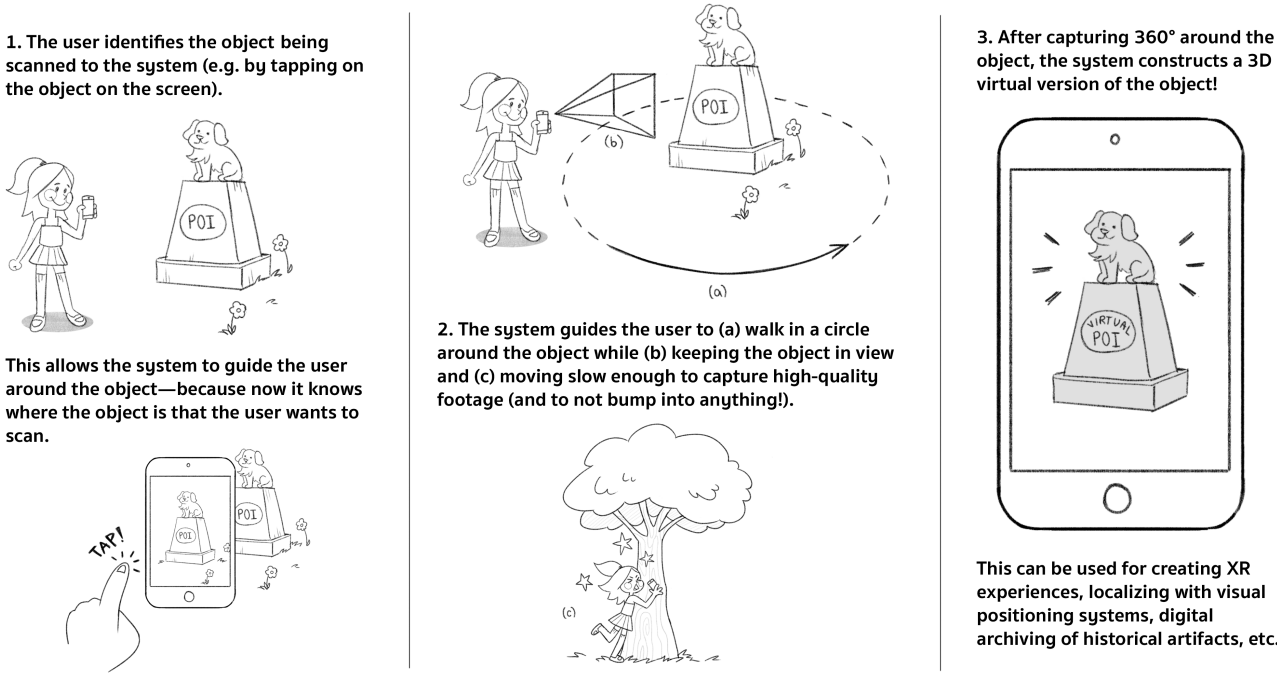


**Figure 1: The basic user flow of our guided 3D scanning apps [101]. We describe a user experience story based on this figure in the supplementary materials. In this paper, we describe (I) the conceptual design of a guided scanning app, (II) an audio/haptic-guided scanning app that we tested in a pilot study, (III) an updated version of our audio/haptic-guided scanning app, and (IV) a visually-guided scanning app, which we compare to III in a final (n=50) user study.**


*CHI '24, May 11–16, 2024, O'ahu, Hawai'i, USA*

## ABSTRACT

People are increasingly using their smartphones to 3D scan objects and landmarks. On one hand, users have intrinsic motivations to scan well, *i.e.* keeping the object in-frame while walking around it to achieve coverage. On the other, users can lose interest when filming


ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/3613904.3642271

inanimate objects, and feel rushed and uncertain of their progress when watching their step in public, seeking to avoid attention.

We set out to guide users while reducing their stress and increasing engagement, by moving away from the on-screen feedback ubiquitous in existing products and apps meant for 3D scanning. Specifically, our novel interface gives users audio/haptic guidance while they scan statue-type landmarks in public. The interface results from a conceptual design process and a pilot study. Ultimately, we tested 50 users in an ultra-high-traffic area of central London. Compared to regular on-screen feedback, users were more engaged, had unchanged stress levels, and produced better scans.

## CCS CONCEPTS

• **Human-centered computing** → *User studies*; **Haptic devices**; **Auditory feedback**; *User interface design*; *Mixed / augmented reality*.

## KEYWORDS

3D scanning, mesh reconstruction accuracy, guidance, multimodal feedback, audio feedback, haptic feedback, visual feedback, AR, engagement, user safety

**ACM Reference Format:**

Jessica Van Brummelen, Liv Piper Urwin, Oliver James Johnston, Mohamed Sayed, and Gabriel Brostow. 2024. Don't Look Now: Audio/Haptic Guidance for 3D Scanning of Landmarks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3613904.3642271

## 1 INTRODUCTION

A smartphone's screen is almost universally accepted as a viewfinder for camera-related tasks. This is fine for casual filming, but systems that provide interactive guidance to the user typically do so by layering extra information on top of the live pixels, *e.g.* [44]. A decade ago, very few non-experts were 3D scanning real-world objects, and [50] worried about how to guide users to capture just 25 images. Scanning guidance tools have barely progressed and largely ignored usability, *e.g.* compare Davis *et al.*'s 3D feedback for people filming to build a LightField [16] versus this year's guidance from Google for capturing NeRFs [91]. At the same time, the user pool has grown massively as people scan things and places for VR purposes [36], 3D printing [82], archaeology [47], medical education [19], and the creation of digital doubles [35, 59], among other use-cases.

Now that iPhones have LIDAR and Android phones compute multi-view stereo in real-time, users are getting increasingly complex, live on-screen feedback about how a scan is progressing. The feedback usually shows a growing mesh or array of feature points that gradually envelops the real-world object if the user walks around it "successfully" [31, 35, 64]. To cope with the known dangers of distracted walking [61], one approach has been to blank the screen [113] so users will look up. We are interested in this kind of safety, though our experiments could not put users at unusual risk, excluding even uneven terrain. We are also interested in engagement, because users would likely scan more objects and could scan them better if they have a pleasant or rewarding experience, compounding their task-specific motivators.

We posit that scanning guidance leveraging modern real-time computer vision could be conveyed more *effectively* using mostly audio and haptics. We define effectiveness relative to an on-screen guidance baseline, and in terms of yielding equivalent 3D scans while potentially improving user engagement and reducing stress levels. We validated this hypothesis in the context of users being guided to scan an outdoor statue in a tourist hotspot.

In terms of overall contributions, we have four main findings. First, our audio/haptic guidance interface was statistically more engaging, usable, and rewarding than a video-only version, as demonstrated on a user study with 50 users. Second, stress levels were not significantly different between the two conditions. Third, the pilot study preceding our main study led us to develop a computer vision based guidance system that significantly reduced users' susceptibility to tracking-drift, which we had found had been a major impediment to guided-scanning. Finally, the scans produced with our audio/haptic guidance were often more accurate than those produced with visual guidance.

## 2 RELATED WORK

We focus here just on works related to scanning, audio and haptic guidance, and geometry reconstruction.

### 2.1 The scanning task

The goal of the scanning task is to generate a virtual, 3D mesh of an area or object using images, video, and/or depth (*e.g.* LiDAR) data [62, 80, 114]. Scanning can vary from rotating handheld objects underneath mounted cameras [80], to taking pictures at optimal locations around objects [92], to methodically walking around while recording video [93, 105].

Using scans, companies, researchers, and other users reconstruct 3D meshes of objects and environments for various downstream tasks. For instance, Google is constructing a 3D map of the world to improve autonomous vehicle navigation [100]. Niantic and Snap are reconstructing points of interest (POIs) with historic or cultural significance for their augmented reality experiences [35, 48]. Archaeologists reconstruct artefacts to better understand history [18, 47].

There are different methods to acquire scan data. For instance, Google sends vehicles with mounted LIDAR and vision systems around the world [100], whereas Niantic crowdsources videos of POIs from Pokémon GO players [48]. Individuals can also create scans for personal projects, such as developing 3D environments of locations dear to them [55], by using various easy-to-access scanning apps [31, 35, 66, 67].

Inspired by the significance of empowering anyone to record locations important to them, as well as how crowdsourcing scans from many individuals can improve reconstruction [48, 95], we focus on the task of individual scanning. Specifically, we develop an app to guide users who may or may not have any prior scanning experience.

Teaching non-experts to scan is no easy task [1, 71, 76], as high-fidelity scans require quality video/images from many different angles around objects [1, 55, 109], and user-friendliness is often

lacking in scanning applications [42]. It is made even more difficult—and even dangerous—when users need to navigate around objects, vehicles, and people, which may be common at locations people want to scan (*e.g.* tourist areas with culturally significant monuments). Thus, we focus on this challenging task: Guiding users to safely walk around a landmark while keeping the object in their camera's view (see Figure 1). We provide a motivational user story of this task in the supplementary materials. To the authors' knowledge, no current applications or research projects address the task of empowering anyone—novice users included—to generate high-quality scans in high-traffic areas.

## 2.2 Audio and haptics for guidance

Scanning can be difficult, so scanning studies and applications use various strategies to guide users in the scanning task. Some instruct users on how to scan before they begin [1, 55]. Others provide users with in-situ visual guidance [1, 37, 74, 80, 109], which researchers have shown to be more effective than prior instruction [1]. In the supplementary materials, we provide in-depth comparisons of various commercial, visual scanning apps. Visual guidance, however, entails walking while looking at a screen and has been shown to be risky [25, 83, 86]. Thus, we investigate whether we could guide users by alternative modalities (*e.g.* audio/haptics) to allow users to look around as they walk.

To our knowledge, no scanning apps have provided scanning guidance solely through audio and haptics—despite this being an effective guidance mode in other contexts. For example, *Shoe-me-the-way* uses haptic vibrations in users' shoes for eyes-free, turn-by-turn navigation [86]. Gallo et al. guided runners using a haptic headband, and found that—compared to turn-by-turn voice-based guidance—the haptic condition provided a better user experience [25]. Others have used haptics to guide users' hands. For example, Ernst and Girouard used haptic-vibration stimuli to teach people to bend objects [20]. Rahman et al. used haptic-guidance to move users' hands toward objects [78]. Multiple studies have used haptics (and/or audio) to guide users to point at a location in 2D or 3D space [10, 27, 58]. Inspired by these works, we employ similar haptic-strategies in our scanning apps to guide users to aim their cameras at POIs.

Guidance technology has also successfully used *audio* for navigation. For example, turn-by-turn voice-based guidance is common in car navigation systems [7]. This high-level, turn-by-turn guidance is great for navigating long distances. However, it does not provide fine-enough detail for guiding people to walk while scanning (*e.g.* keeping users at meter-level distances as they circle POIs) [7, 78]. Researchers have developed alternative audio-guidance systems for this reason. For example, Menelas et al. used impact sounds, which varied in frequency with respect to distance, to guide users' aim [58]. Others use spatial audio to allow users to pinpoint locations in space [27, 56, 102]. Others have utilized variations in music (*e.g.* pitch, volume) to provide navigation guidance [25, 40]. Like these examples, we also use music in our scanning app, as it is both effective and enjoyable for navigation guidance [25, 40].

## 2.3 Reconstructing geometry from scans

3D reconstruction remains a core task in the Computer Vision community, with excellent surveys [103] and tutorials [24]. Algorithms keep advancing to cope with different hardware, software, and situational settings. In the supplemental materials we walk through the considerations relevant to smartphone-based scanning. To summarize here for brevity: the most accurate geometry is obtained with accurate camera poses, sufficient coverage of the subject with *many* overlapping views (a 360° loop is preferred), blur-free images, and good camera baselines. The camera-to-subject distance should be constant and small, though the ideal radius depends on the height of the object, and the sensor range if using a LIDAR phone. Unlike the final accuracy-focused reconstructions, the phone's UI requires real-time processing (see Figure 3 for examples).

# 3 STAGE 1: INITIAL SCANNING APP DESIGN

Learning to scan well is not a simple task, and non-expert users typically need study facilitators or in-app feedback to help [1, 55, 71, 109]. Because of this learning curve, we designed our guided-scanning app using Jackson's Theory of Conceptual Design [38, 39]. Jackson's theory argues that by viewing designs in terms of *concepts*—or the fundamental ideas users need to understand to use a system—designers can create straightforward, easy-to-learn systems [38, 39]. In the supplementary materials, we present the details of our conceptual design. Table 1's first column explains the concepts core to our design: "C1: *Object Transform*", "C2: *Scan*", and "C3: *Mesh Reconstruction*".

Additionally, for our design, we adopt the well-tested strategy of providing in-app, immediate feedback to guide users as they scan [1, 74, 80, 109]. Because reconstructing POIs with sufficient detail requires slowly capturing images from many different angles, and at distances where texture detail is still visible [1, 43, 55, 109], we included the following guidance features in our design: "F1: Distance guidance", "F2: Framing guidance", "F3: Speed guidance", and "F4: Completion guidance" (see Table 1). The Video Figure in the supplementary materials illustrates these features.

# 4 STAGE 2: EXPLORATORY PILOT STUDY

The main goals for this stage of the scanning app design were to determine whether non-expert users could be guided via audio/haptics to scan a POI outdoors in a busy area, and to obtain qualitative feedback on the initial audio/haptic prototype. Specifically, we wanted to investigate what users enjoyed about the scanning process, what caused them stress, and whether they felt more aware of their surroundings using an audio/haptic-based app (*e.g.* because they no longer needed to look at a screen).

For this study, we developed an initial audio/haptic scanning prototype app, as described in Section 4.1. We compared this prototype to a well-established visual scanning app, "Wayfarer" [3, 48], in a counterbalanced, within-subjects study. The results of this study heavily influenced our final scanning app design. For the final design, see Section 5.1.

Table 1: The progression of our scanning guidance apps in terms of concepts (C) and features (F). The first column describes our initial conceptual design, the second describes the pilot study apps, and the third describes our full-scale study apps. Blue indicates an audio/haptic-guided app, and magenta indicates a visually-guided app.

| | Stage 1: **Initial Design** | Stage 2: **Pilot Study Audio/Haptic Prototype (Ours)** and **Visual Baseline (Wayfarer [66])** | Stage 3: **Full-Scale Study Audio/Haptic App (Ours)** and **Visual Baseline (Ours)** |
|---|---|---|---|
| Concepts | **C1:** *Object Transform*: For the system to guide the user, it needs to know what/where the user wants to scan: The user must first convey to the system the physical POI's location, orientation, and scale. Then the system can guide the scanning process. We call the virtual position/orientation/scale of this object the "Object Transform". | **Pilot Audio/Haptic Prototype**: To communicate the Object Transform concept to the user, our prototype asks the user to "tell the computer where the object [they] want to scan is" by aligning a virtual placeholder object to the physical object's location. The virtual placeholder object is an AR translucent marble statue, as shown in Figure 2. We initially chose to represent the transform as a statue because people often scan statues as points of interest [63], and in our study, users scan a statue [107].<br>**Pilot Visual Baseline:** The Wayfarer app does not provide scanning guidance to the user. and therefore does not include the *Object Transform* concept. | Both **Audio/Haptic App** and **Visual Baseline**: These apps communicate the *Object Transform* concept similarly to the pilot audio/haptic prototype, with the following upgrades:<br>• We changed the affordance for the *Object Transform* to be a 3D bounding-box because, in the pilot study, the statue representation was confusing for some users<br>• We used an interactive image segmentation model to allow users to simply tap the object they wanted to scan (instead of having to directly manipulate the statue transform) |
| | **C2:** *Scan*: 3D reconstruction depends on obtaining images and pose information from many different angles around the POI, to faithfully compute the shape and texture of a virtual doppelganger. We call the images and pose information "the scan", and the user's process of moving around the object, "scanning". | **Pilot Audio/Haptic Prototype:** Our prototype uses a scanning tutorial to communicate the concept of *Scanning* to the user. In this tutorial, users get to experience the various guidance mechanisms (*i.e.* distance/framing/speed/completion guidance, described below), which teach them how to scan well.<br>**Pilot Visual Baseline:** The Wayfarer app partially communicates the *Scanning* concept to the user through a recording button at the bottom of the screen; however, the user should have prior knowledge about how to scan before using the app. | Both **Audio/Haptic App** and **Visual Baseline**: These apps communicate *Scanning* to the user the same way the pilot audio/haptic prototype does, with the difference that we changed the guidance mechanisms, as described below. |
| | **C3:** *Mesh Reconstruction*: The purpose of scanning is to develop virtual version of POIs. We call this virtual 3D version the "mesh reconstruction". | **Pilot Audio/Haptic Prototype:** Our prototype locally generates low-fidelity mesh reconstructions on the phone after users successfully complete scans. The reconstruction takes approximately 10-45 seconds. Users can explore the mesh by pinching to scale and swiping to rotate. Figure 3 shows example meshes generated by the app.<br>**Pilot Visual Baseline:** Not implemented. | Both **Audio/Haptic App** and **Visual Baseline**: These apps generate *Mesh Reconstructions* the same way as the pilot audio/haptic prototype does. |
| Guidance Features | **F1:** Distance guidance: To scan well, the user should walk in a ring around the POI, close enough to capture its detail, and far enough to obtain a comprehensive view. The system can guide the user to stay at such a distance. | **Pilot Audio/Haptic Prototype:** If the user is at the correct distance away from the *Object Transform*, our app plays music. If they are too close or far away, the volume of the music decreases.<br>**Pilot Visual Baseline:** Not implemented. | **Audio/Haptic App:** This app implements distance guidance similarly to the pilot audio/haptic prototype, with the following addition: When the user is too close, the music's pitch/speed increases, and when the user is too far, the music's pitch/speed decreases.<br>**Visual Baseline:** An AR blue track appears around the *Object Transform*, indicating the correct distance at which to walk around the object (see Figure 8). The tutorial tells the user to walk along this blue track. |
| | **F2:** Framing guidance: To scan well, the user needs to keep the POI in frame. The system can monitor the camera angle and guide the user's aim. | **Pilot Audio/Haptic Prototype:** When the *Object Transform* is not in view, haptic vibrations from the phone pulse at a steady rate. Additionally, a beeping noise plays. As the camera angle deviates further from the *Object Transform*, the beeping noise increasingly reverberates.<br>**Pilot Visual Baseline:** Not implemented. | **Audio/Haptic App:** When the *Object Transform* is not in view, the user experiences haptic feedback. The feedback increases in frequency and intensity as the camera angle deviates further from the *Object Transform*. There is no audio feedback associated with framing guidance in this version (as users found the beeping noise in the pilot study to interfere with the distance-guidance audio).<br>**Visual Baseline:** When the *Object Transform* is not in view, translucent arrows appear on screen (see Figure 8). The opacity of the arrows increases as the camera angle deviates further from the *Object Transform*. |
| | **F3:** Speed guidance: To scan well, the user needs to maintain a slow pace. The system can guide the user to slow down, if needed. | **Pilot Audio/Haptic Prototype:** When the user moves at an angular speed at which the camera cannot capture enough frames, an audible voice warns the user (*e.g.* by saying "Too fast").<br>**Pilot Visual Baseline:** The text, "Slow Down", appears on screen. | **Audio/Haptic App:** This app implements speed guidance in the same way as the pilot audio/haptic prototype does.<br>**Visual Baseline:** The text, "Slow Down", appears on screen. |
| | **F4:** Completion guidance: To scan well, the user needs to capture the POI from many different perspectives. The system can let the user know when they have provided the system with footage from all sides and thus "completed" the scan. | **Pilot Audio/Haptic Prototype:** When the user walks 360° around the *Object Transform*, an audible voice tells the user they completed the scan (*e.g.* by saying, "You're done!").<br>**Pilot Visual Baseline:** Not implemented (although a message appears upon starting to scan, "Scans must be longer than 20 secs"). | **Audio/Haptic App:** This app implements completion guidance in the same way that the pilot audio/haptic prototype does.<br>**Visual Baseline:** When the user walks 360° around the *Object Transform*, a screen with the text, "Processing your scan", appears. |
| Other Features | - | - | **F5:** Drift-reducing algorithm: Both **Audio/Haptic App** and **Visual Baseline**: In the pilot study, users mentioned drift being a major stressor. Thus, we developed a new drift-reducing algorithm, "Landmark Tracker" (as described in Section 5.1), which we implemented in both of these apps. This significantly reduced drift. |
| | - | - | **F6:** Grip improvement: Both **Audio/Haptic App** and **Visual Baseline**: In the full-study, we provided users with phone grippers, as users in the pilot study mentioned holding the phone openly (while considering potential phone thieves) was stressful. |
| | - | - | **F7:** Sound-reduction in public: **Audio/Haptic App**: We provided users with hear-through headphones, as users in the pilot study mentioned it was stressful having loud app sounds in public. |

## 4.1 Audio/haptic scanning app implementation (Pilot Version)

The pilot prototype of the audio/haptic scanning app is described in Column 2 of Table 1, with respect to the app's concepts and features. We used Unity [97] and Lightship ARDK v2.5.2 [65], which in turn uses ARKit for tracking pose [53], to develop the scanning app. We also used the Unity Taptic Plugin [32], Stable Diffusion [70] to develop icons, and an audio file from Freesound [99].

## 4.2 Visual scanning app for comparison (Pilot Version)

To determine whether our audio/haptic scanning app was comparable to a current, visual scanning app, we had users try Wayfarer, a scanning app commonly used in various gaming communities (*e.g.* Pokémon GO, Ingress, *etc.* [3, 48, 64, 66]), as an exploratory baseline. We specifically chose Wayfarer because the code for Wayfarer's scanning capabilities is freely available through Lightship ARDK [65].

As shown in Figure 4, Wayfarer provides visual feedback to the user in the form of a dark overlay on the camera view, which slowly disappears as the scan progresses. It also provides feedback on the users' speed by displaying text on screen ("Slow Down") when the user is moving too quickly. Note that Wayfarer does not implement all of our concepts or guidance features, as shown in Table 1. However, we develop a visual app that *does* in Section 5.2 for the full-scale study.

## 4.3 Pilot study procedure, participants, and measures

Before scanning the statue, participants (n=6) completed a Research Consent Form, obtained an anonymous codename, and watched an introductory video, which provided an overview of the study and described the concepts of *AR* and *POIs*. We then gave them iPhone 13 Pros, and they completed two practice scans indoors: once using our audio/haptic app and a second time using the visual baseline app (or in reverse order when counterbalancing). They could use the apps for as long as they wanted.

Next, they went outdoors to the Agatha Christie Memorial statue [107], and completed a scan of the statue using either the audio/haptic app or the visual app (depending on counterbalancing). The scanning task involved walking around the statue 360° while aiming the phone's camera at the statue (see Figure 1). They then repeated this process with the other app, and returned indoors for a semi-structured interview, where we asked them to compare the visual and audio/haptic scanning apps. E.g. we asked participants, "Was there anything particularly stressful in either scanning experience?". We provide the interview questions and additional participant demographics in the supplementary materials.

The location of the study was chosen due to the cultural significance of the statue [79, 107], and the busyness of the area (which is next to the Leicester Square station entrance [22]). This provided participants with an adequately challenging scanning task, as described in Section 2.1. Each participant completed the study individually at separate appointments, and the condition order (audio/haptic *vs.* visual baseline) was counterbalanced (see Table 2).

**Table 2: The pilot study procedure, which took approximately 45-60 minutes per participant. We switched the order of (A) and (B) when counterbalancing the study.**

| Activity | Estimated length (minutes) |
|---|---|
| Introductory Video | 5 |
| Indoor Scanning Practice | 15 |
| A. Visual App: Outdoor Scan | 4 |
| B. Audio/Haptic App: Outdoor Scan | 4 |
| Semi-structured Interview | 15 |
| Demographics Survey | 2 |
| **Overall** | **45-60 minutes** |

## 4.4 Pilot study analysis

The pilot study was within-subjects and counterbalanced by swapping the visual baseline and audio/haptic scanning order. We collected qualitative data about participants' scanning experiences in semi-structured interviews. The results from the pilot study are exploratory, and provided us with information about how to improve the app design before starting the full study. We present them here, as exploratory research can provide key insights to others designing in similar fields [104, 108].

To analyse the interview data, we performed a reflexive thematic analysis according to the methods described in [5], with the caveat that a single researcher performed the majority of the coding process, with a second researcher observing and commenting on the themes (hence why we describe these results as exploratory). Through the six-phase, iterative process of theme identification [5], themes emerged with respect to stressors, stress reducers, rewarding aspects of the experience, and opportunities to improve the app.

## 4.5 Pilot study results

Figure 5 shows the themes from our qualitative analysis. In the following sections, we describe how the major themes shaped our full-study scanning app design. We walk through each of the themes with user quotations and detailed analysis in the supplementary materials. Note that feedback in the Visual app was subtle enough to conceal moderate drift if it occurred. Below, we provide a brief overview.

*4.5.1 Stressors and stress reducers.* In this study, we were interested in what users found difficult or caused them stress, so that we could minimize these in the next iteration of the app. All six participants mentioned one or more stressors, and various aspects of the apps that reduced their stress, or ideas for reducing their stress. We organized these into the three categories outlined in [87]: Social-evaluative, cognitive, and physical stress.

In terms of social-evaluative stress, five of six participants mentioned stressors involving other people's perceptions of them. E.g. P3 and P4 mentioned the visual scanning condition felt invasive of people's privacy because it was obvious they were recording (P3: "For the video app, I saw people duck away a couple of times [...]

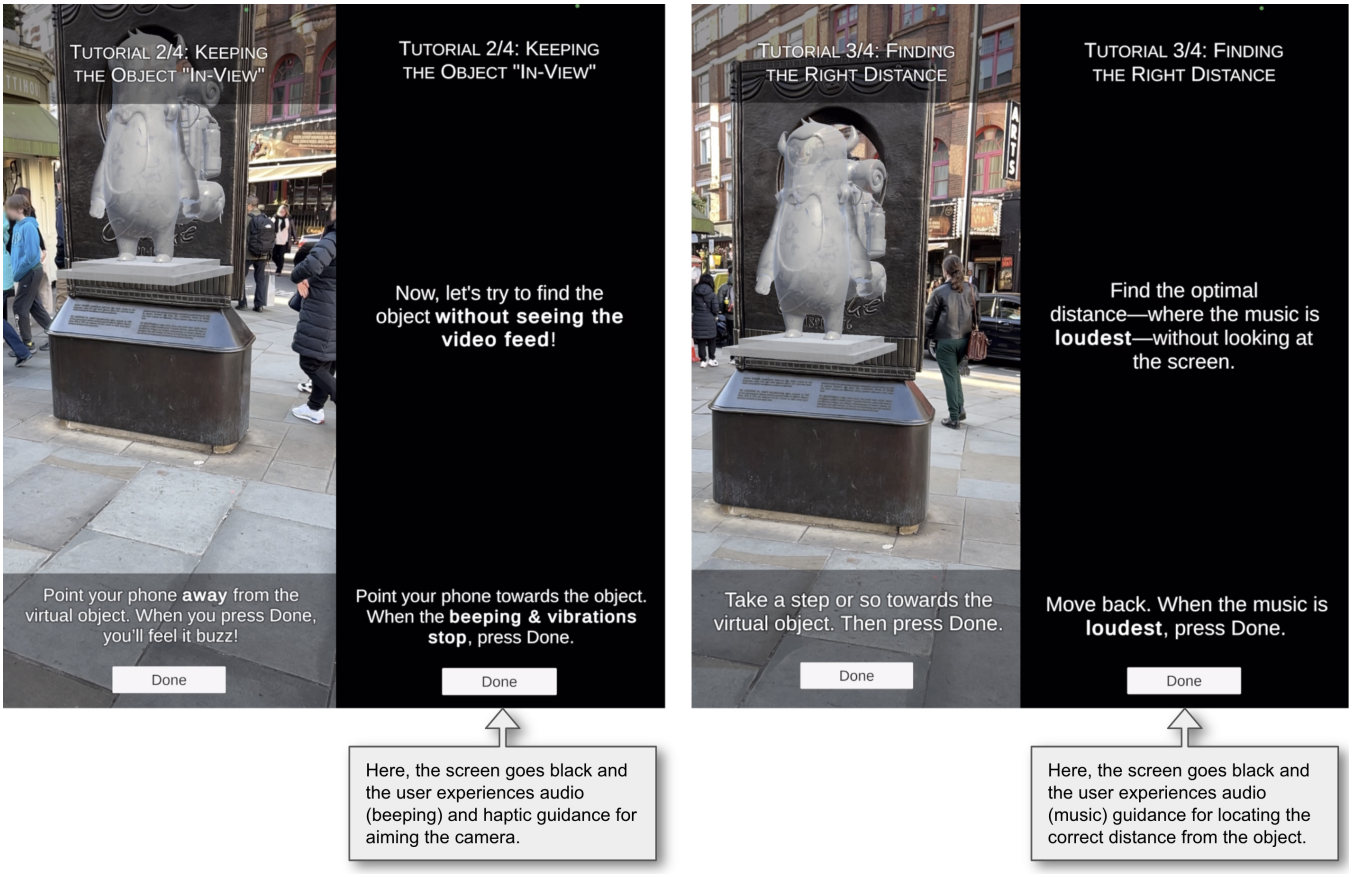


**Figure 2: Screenshots from two of the tutorial steps in the pilot study version of the audio/haptic scanning app. The virtual marble yeti statue represents the *Object Transform* of the POI, which users place on top of the real POI (a rectangular statue, in this case). The position/orientation/scale of the *Object Transform* informs the system about the scene layout, so it can guide the user as they scan.**

because they really thought I was videoing. [...] The [audio/haptic app] was a little bit more socially acceptable, mainly because if [bystanders] happened to look at my screen, it didn't seem like I was recording"). Four participants felt the loud app sounds when using the audio/haptic app in a public space were stressful. Thus, for the final study, we incorporated headphones, and recommend **minimizing external audio in public** (F7 in Table 1).

There were two main cognitive stressors with the audio/haptic app. The first was encountering a bug in which users would not see the completion screen after finishing their scan. Three participants mentioned encountering this. The second was feeling as if the audio/haptic feedback was not aligning with the location of the actual POI. E.g. P2 described how the system was providing a lot of negative feedback ("buzzing"), which "could have been because it shifted place in the middle [of the scan] or something". We suspected this was because the AR session had lost track of the environment (*e.g.* due to SLAM drift [98]), and the *Object Transform* had drifted away from the physical POI location. Thus, we decided we needed to update our app by **minimizing drift and eliminating any bugs** (F5 in Table 1).

In terms of the visual app, two users found the lack of guidance stressful. E.g. P1 said, "The audio one was way more like, 'Good job, you're doing the right thing!', but with the [visual app] sometimes I couldn't see the object". Thus, we recommend **providing scanning guidance to increase feelings of confidence**. Nonetheless, two other participants felt their cognitive load decreased with the visual app because they had prior experience with similar apps.

In terms of physical stress, five of six participants mentioned they felt like they may collide with something, especially when looking at the phone screen. For example, P4 stated, "The video [app] makes you more focused and you're less aware of your surroundings. [...] With the video [app], I bumped into somebody just because I had to focus on the camera". Participants also described how—although they felt people may steal from them while scanning—using the audio/haptic app could allow them to "look around more [...] for pickpockets". Nonetheless, two participants mentioned they felt like they needed to grip the phone more strongly with the audio/haptic app because they were not looking at it. Because of this, we provided participants with phone grippers in the full-scale study, and recommend **improving users' grip, when possible** (F6


**Figure 3: Our scanning app reconstructed these meshes from users' scans.**

in Table 1). We also recommend **hiding the video feed to increase users' feelings of awareness and reduce social-evaluative stress**.

*4.5.2 Rewarding aspects and other opportunities.* Users mentioned various rewarding aspects and opportunities for the audio/haptic app. We present the following recommendations—which are directly based on user sentiments, as described in the supplementary materials:

- **Use music and voiced feedback as rewarding features**,
- **Provide an in-app mesh reconstruction (even low-fidelity) as a rewarding feature**,
- **Utilize haptic feedback for view-finding as a rewarding feature**,
- **Better convey the *Object Transform* concept by using a neutral virtual object (*e.g.* bounding-box), which fully encapsulates the physical object** (C1 in Table 1), and
- **Utilize tone changes to improve music guidance** (F1 in Table 1).

### 4.6 Pilot study conclusions

From the interviews, we noticed the majority of the stress-related comments about the audio/haptic app (81.3%) were regarding the "loud app sounds" (43.8%) and "encountering a bug/drift" (37.8%), as shown in Figure 5. Thus, we hypothesized that by removing any bugs and app sounds in public, and reducing drift, we could significantly improve the user experience. We summarize our other updates to the app in Table 1. Other key results included how many users felt the guidance boosted their scanning confidence, and found the audio/haptic features rewarding.

## 5 STAGE 3: FULL-SCALE STUDY

Now, the pilot study's app itself was overhauled, improving on most of the Concepts and Features. Please see the last column of Table 1, which summarizes the two alternatives implemented in the app for the final study, namely the **Audio/Haptic App** *vs.* the **Visual Baseline**. This new and final split-mode app allowed us to complete a full-scale (n=50) study. It sought to validate our hypothesis (from Section 1) that scanning guidance leveraging modern computer

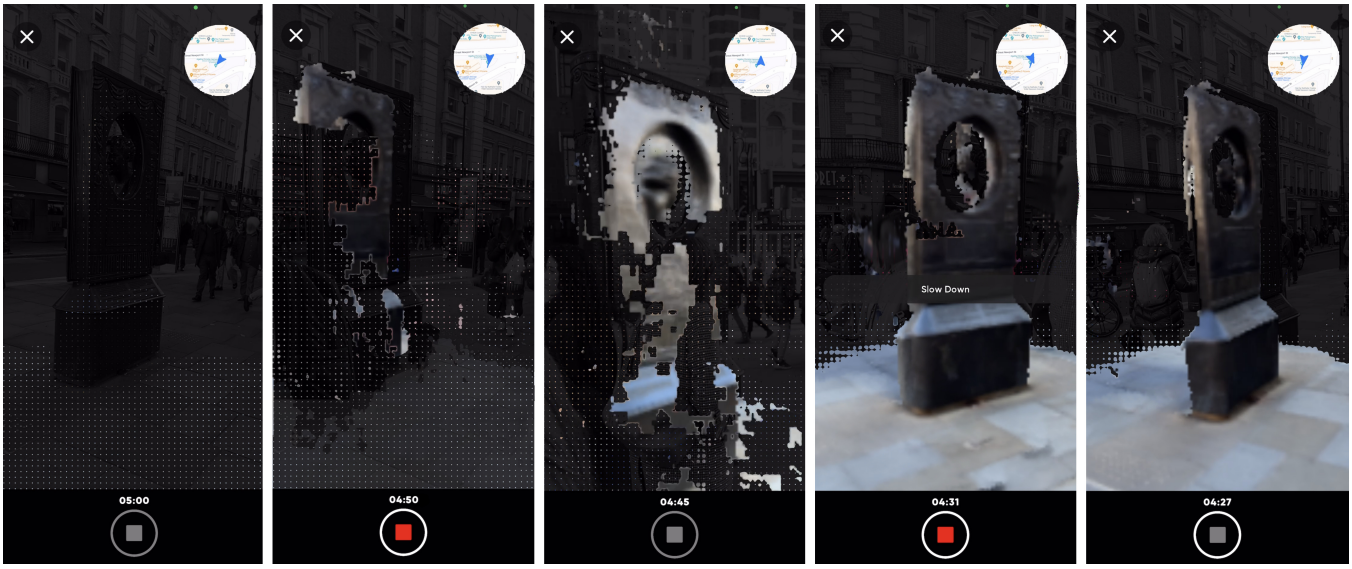


**Figure 4: Screenshots from the Wayfarer app [66]. Notice how the dark overlay over the camera view disappears in certain areas as the scan progresses. Also notice the "Slow Down" notification in the fourth screenshot.**

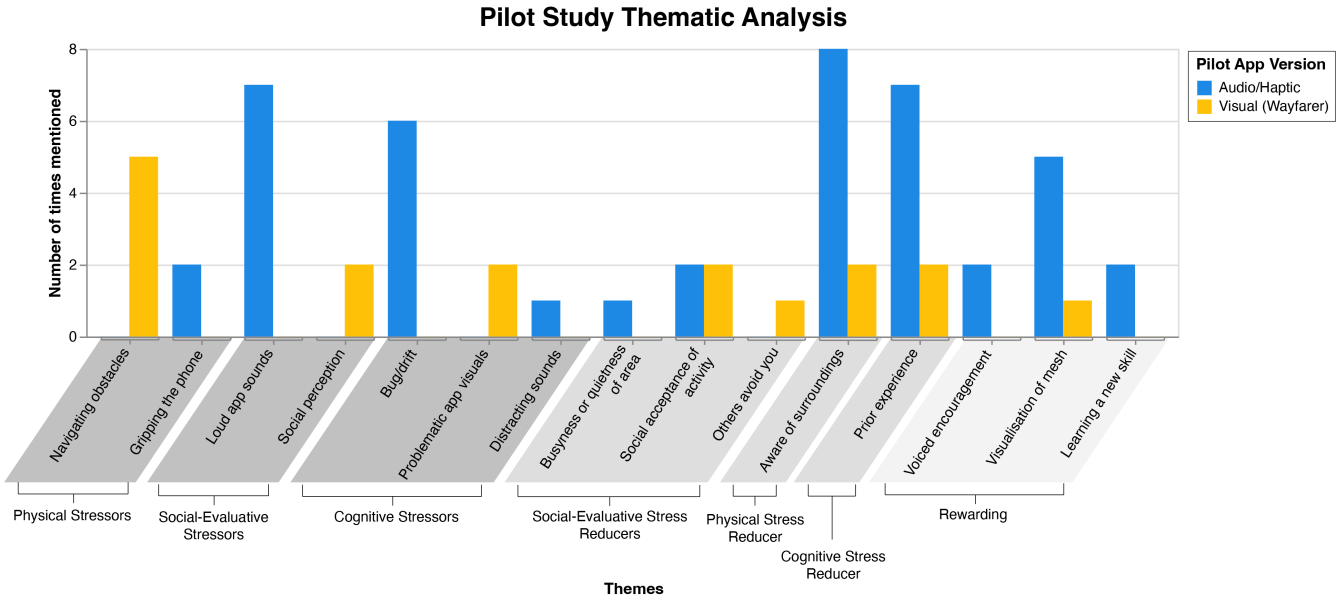


**Figure 5: The number of times participants mentioned various themes in the semi-structured interviews. Stressors are highlighted in dark-grey; stress-reducers in medium-grey; and rewarding features in light-grey. Note how the main stressors in the audio/haptic condition were the loud app sounds and encountering a bug/drift. In terms of stress reducers in the audio/haptic condition, participants felt aware of their surroundings, and wished they had more prior experience with audio/haptics.**

vision could be conveyed more effectively using mostly audio and haptics.

For fairness, the two conditions need to be comparable. In the pilot study, participants appreciated the scanning guidance provided by the audio/haptic app. It was possible that users were responding to having guidance of *any kind*. So, the final app includes all the same algorithmic and interface improvements that could be shared, but one version provides visual guidance, where the other uses audio/haptics. The shared improvements, *e.g.* for drift-reduction, are described next, but we note that consequently, the version with visual guidance is a kind of "super-baseline" compared to existing

apps and systems (see the supplementary materials for a detailed comparison).

Drift was a significant stressor in the pilot study, despite using ARKit [53], which includes Apple's proprietary tracking software and is well-used in the literature [57, 84, 96, 110, 112]. Drift can emerge when the phone's pose-tracking vision system loses sight of too many static visual features in an environment [52]. This happened to five of six users at some point, because of filming outdoors and because of people occluding parts of the scene, despite the phone's sophisticated ARKit library which uses visual and inertial SLAM [2]. We address **drift-reduction** in Sec 5.1.

In addition to noticing drift in the *pilot* interviews, we noticed certain aspects of scanning can be rewarding, as well as stressful, and users can find it difficult to know how to create accurate meshes. Thus, we compared our audio/haptic and visual scanning guidance (given the task of scanning the Agatha Christie Memorial) with respect to the following metrics:

- **engagement** (including rewarding factors, user attention, and usability [73]),
- **stress**,
- **safety**,
- **user accuracy** in following the guidance, and
- **mesh reconstruction accuracy**.

## 5.1 Changes to the audio/haptic scanning app

In the pilot, users had substantial difficulty understanding and placing the *Object Transform* proxy-shape (shown in Figure 2), despite proxy-cubes with handles being used in previous apps, and in the upcoming Apple Object Capture module [34]. Ignoring UI aspects, such a computer-vision assisted initialization process made sense until now, because pose-estimation/SLAM has been fast enough to run real-time on edge-devices since [17].

We set out to ameliorate drift and improve the initialization UI by changing what the user communicates to the phone at the outset. Instead of specifying *where* the POI sits in the phone's somewhat fragile coordinate system, the user picks out what the POI looks like. For this new approach, we added an interactive image segmentation CNN ("MagicTouch" by Google [4]) to the Unity app, via TensorFlow Lite for Unity [33]. We also built a 3D bounding volume estimation algorithm (described in the supplementary materials) based on the output pixels from the segmentation model. With these changes, the user no longer needed to manually manipulate the position/orientation/scale of the *Object Transform*. They merely needed to tap on the object to make a bounding-box appear, as shown in Figure 6.

Further, to overcome mid-scan drift (F5 in Table 1), we made a custom "Landmark Tracker" algorithm, shown in Listing 1 and illustrated in Figure 7. It already has the POI's current appearance from the user's initialization. As the user walks around, both the pose and appearance will change. Therefore, whenever possible, the system re-estimates the bounding-box transform using the interactive segmentation model [4]. However, the segmentation model's seed-point is no longer the *user's* input. Instead, it is the projection of the centerpoint of the previous reliable bounding-box onto the camera image plane. Using this point, the object is once again segmented and 3D bounding-box is computed. This bounding-box is Kalman Filtered [106], and the process repeats until the user completes the scan.

**Listing 1: Pseudo-code of the drift-reducing algorithm, "Landmark Tracker"**

```
1. User taps on-screen to identify the object to scan
    a. Segment the object pixels using the location (
 uv) of the user's tap
    b. Based on the object pixels, estimate a 3D
 bounding-box (i.e. "object transform" with position/
 orientation/scale)
    c. Use this transform as the first input to an 9
 DOF Kalman filter
2. Project the center of the bounding-box (xyz) onto
 the camera image plane (uv)
3. IF this location (uv) is on-screen (i.e. the
 bounding-box is in-view),
    a. Use the location on-screen to as input to the
 interactive segmenter
    b. Estimate a new object transform
ELSE return to Step 2
4. IF the user is not looking at a short edge of the
 bounding-box (i.e. len(front edge) >> len(
 perpendicular edge) ),
    a. Update the Kalman filter model using this box
 as input
ELSE return to Step 2
5. IF the output of the Kalman filter is not
 significantly different from the previous transform
 (i.e.  abs((original transform) - (new transform)) <
  epsilon),
    a. Replace the original bounding-box with this
 new estimate
6. Repeat from Step 2
```

The other challenges included refining the distance and framing guidance. To upgrade the distance guidance, we dynamically modified the music. Its pitch and speed increased gradually as the user moved too close to the object, and decreased if they walked too far away. This allowed users to orbit the object based on audio. To keep the audio private to the user, we provided them with "hear-through" headphones (Jabra Elite 5 [29]), so they could both hear outside noise (*e.g.* nearby cars), and the app's audio.

We upgraded the framing guidance which comes across through haptic vibrations in the phone. In the pilot version, beeping sounds and steady-rate phone pulses notified the user if the POI went out of frame. Now, as the camera angle deviates further from the *Object Transform*, the haptic feedback increases in frequency and intensity. The audio feedback was discontinued here because users found the beeping noise in the pilot interfered with the distance-guidance audio.

We also provided users with phone grippers with the aim of reducing stress related to phone-theft (see F6 in Table 1).

## 5.2 Visual scanning app implementation

Our visually-guided scanning app implemented all of the core "concepts" and "features" of the audio/haptic scanning app, as described in Table 1. Thus, for the full-study, we could directly compare audio/haptic and visual scanning guidance. We also implemented green mesh visualizations, similar to Snap's scanning app [35], as shown in Figure 8.


**Figure 6: The full-study apps use our image segmentation with 3D bounding-box estimation to place the *Object Transform*. The user sees only the first and third screen.**

Note that although it may seem like the user has to point the camera away from the *Object Transform* to view the distance guidance (*i.e.* the blue track around the object in Figure 8), if the user is in the middle of the track and aims the camera at the statue, the track is generally visible due to its scale, its height above-ground (which is adjusted based on the camera's location above-ground), and the camera's field-of-view. Additionally, we show there were no significant differences between the percent time users kept the *Object Transform* on-screen between conditions in the supplemental materials.

## 5.3 Study procedure

The full-scale study procedure (see Table 3) was nearly identical to the pilot procedure. However, we employed a between-subjects design for the full-scale study to minimize learning and fatigue effects. Thus, participants either used our audio/haptic *or* visual app in the study. The participants did not know about the other app design until they completed the study, when we gave them the opportunity to try it. Other differences in the procedures included the surveys and measures.

*5.3.1 Surveys and measures.* Participants completed short pre- and post-surveys before/after scanning, and a final survey once indoors.

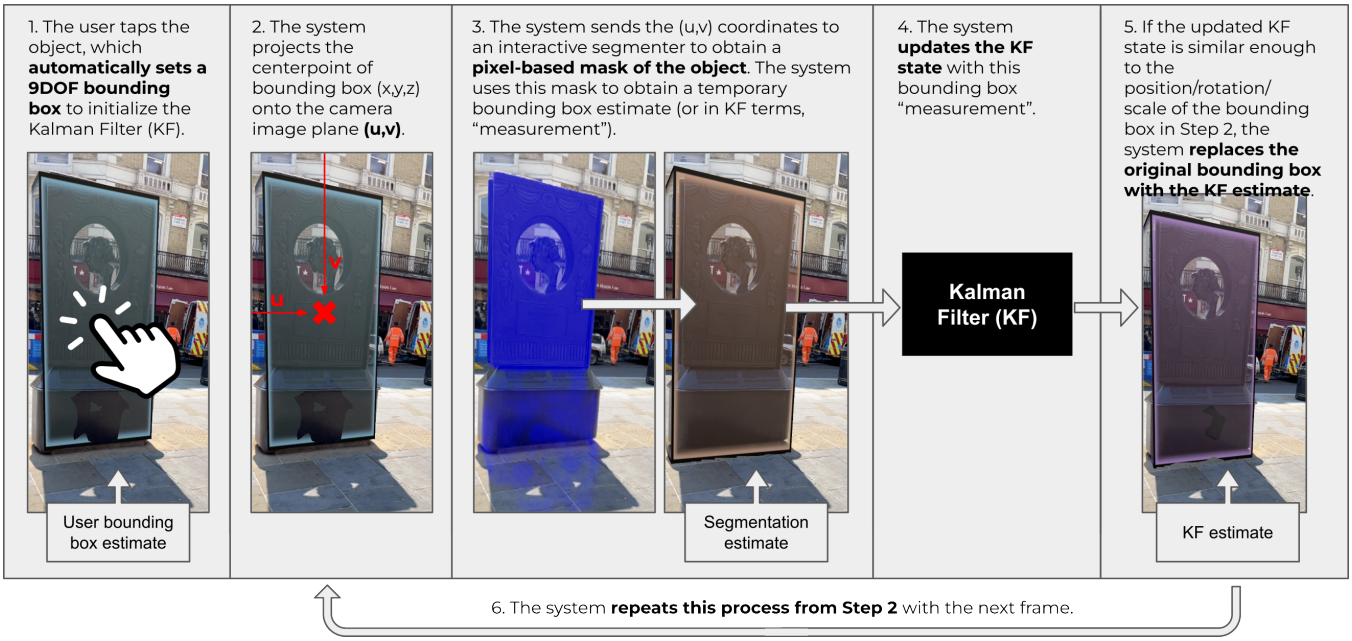


**Figure 7: A visualization of the Landmark Tracker algorithm, which we developed to reduce drift of the *Object Transform* in the full-study apps.**

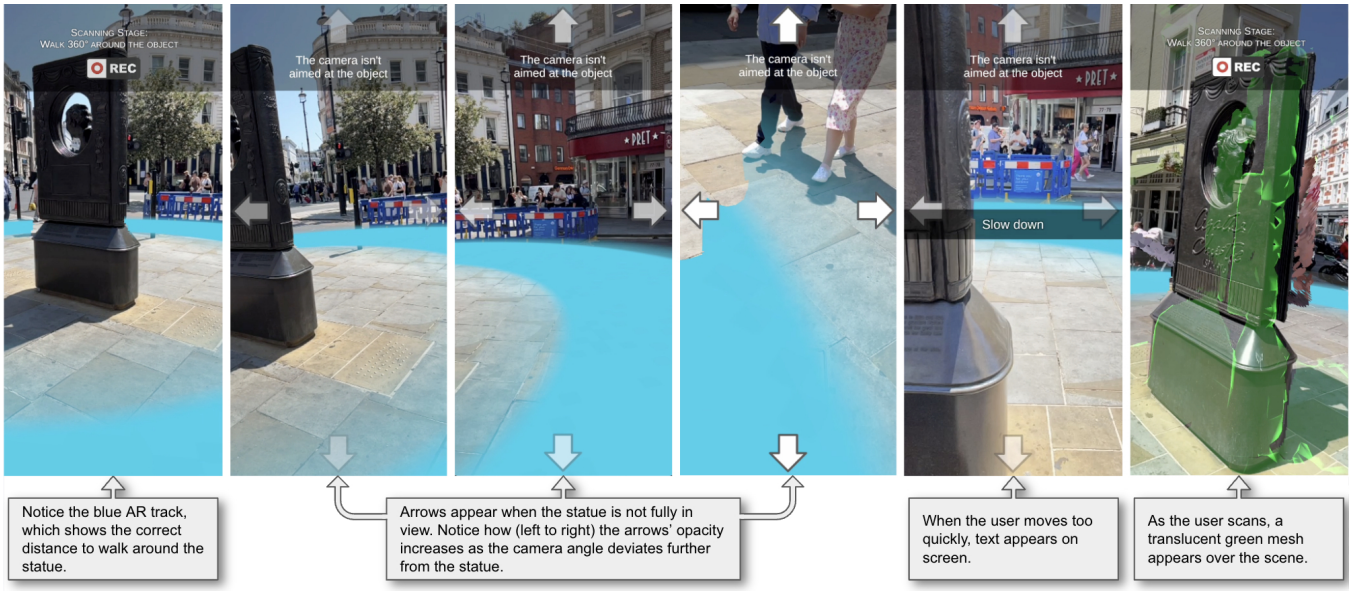


**Figure 8: Screenshots from our visual scanning app. This version was used for the full study.**

These surveys are in the supplementary materials and are summarized in Table 4.

As shown, the surveys included well-established questionnaires, like the State-Trait Anxiety Inventory for determining whether someone is characteristically anxious [94], and questions we developed based on the pilot study. For instance, we asked participants to check boxes containing anything that contributed to their stress levels. The boxes contained stressors mentioned in the pilot study, including "navigating around people", "thieves", *etc.* We presented the survey questions in randomized order.

To measure engagement, we used the UES-SF. The questionnaire's basic form includes questions about aesthetic appeal. However, the two app modalities' aesthetics (audio/haptic *vs.* visual) are

**Table 3: The full-scale study procedure. Each participant used only one of either our audio/haptic or our visual scanning app.**

| Activity | Estimated length (minutes) |
|---|---|
| Introductory Video | 5 |
| Indoor Scanning Practice | 10 |
| Slow Outdoor Walk for Baseline Heart Rate | 1 |
| Short Pre-Survey | 2 |
| Outdoor Scan of the Agatha Christie Memorial [107] | 5 |
| Short Post-Survey | 2 |
| Final Survey | 15 |
| **Overall** | **40-60 minutes** |

**Table 4: Summary of the measures in the full-scale study surveys.**

| Survey | Measure | Scale |
|---|---|---|
| Pre- and post-survey | Stress categories: Social-evaluative, cognitive, physical [87] | Self-reports with Likert Scale [28, 30] |
| | Emotional affect | Self-Assessment Manikins [60] |
| Post-survey | Physical safety proxy | Self-report of number of collisions during the scan |
| Final survey | Experience scanning | Prior knowledge about scanning (multiple choice) |
| | | Number of prior scans completed |
| | | Any profession/hobbies related to scanning (short answer) |
| | Gender (heart rate-related) | Short answer [81, 85] |
| | Age (heart rate-related) | Number [81] |
| | Anxiety characteristic | State-Trait Anxiety Inventory [94] |
| | Particular stressors (identified in the pilot study) | Checkboxes for applicable stressors |
| | Engagement: Focused attention, reward factor, perceived usability | User Engagement Scale - Short Form (UES-SF) [73] |

quite different, so we chose to only include questions with respect to focused attention, reward factor, and perceived usability [73].

We also measured participants' heart rate (HR) as a proxy for stress [77, 87], as heart rate does not suffer from the subjectivity of self-reports [77, 87]. We used a wrist monitor, the Garmin Venu 2 Plus [54], as it is less invasive than chest-worn devices [87], but can provide similar accuracy/precision [12, 23, 111]. We measured participants' HR during a 30+ second baseline walk around the statue, and during the scanning task. This allowed us to calculate their incremental HR, *i.e.* the difference between their average baseline HR and their average experimental (scanning) HR [8, 49, 51].

We measured a number of potential confounding factors and covariates, including participants' gender, age, and anxiety characteristic, which could affect HR; and their prior experience scanning, which could affect scanning ability. We also speculated that the busyness of the area could contribute to participants' stress. Thus, we incorporated proxy variable for busyness in the analysis. We obtained this proxy, a "busyness score", using the TFL Crowding API [22] at the transit stop, "Leicester Square". The transit stop is very close to the Agatha Christie Statue [22, 107].

The final measures are related to users' accuracy, and the accuracy of their mesh reconstructions. Using in-app logs, we recorded the time users spent outside of the "correct distance" (*e.g.* the visual app's "blue track" in Figure 8). We also recorded the time the *Object Transform* was not in view, the number of speed warnings (see Table 1), and the total length of the scan.

We also obtained LIDAR depths and images recorded by the iPhone 13 Pro during the scan. (Note that we blurred recorded faces for anonymization.) As described later, we used this data to reconstruct meshes and compute scan accuracy scores.

## 5.4 Participants

Fifty participants completed the study in London. On arrival, participants signed a research consent form and were provided with an anonymous codename. Recruited participants' ages ranged from 18 to 53 ($\bar{x}$=29.66, SD=9.07) with 38% of participants being female, 58% being male and 4% being non-binary or transgender.

Prior experience with 3D scanning was not a requirement for recruits. Instead, the study was widely advertised through various social media user-study groups; Niantic and University College London (UCL) social channels; word-of-mouth on the street; and posters at UCL and near the study location. We asked interested individuals to complete an availability form and then contacted them to confirm an appointment. By the end of the study, this form had received over 160 applications, which resulted in 50 sessions. Each participant received a £25 gift card and complimentary merchandise (*e.g.* stickers).

## 5.5 Data cleaning

Not all participants completed each survey or the required HR measurements. Six participants forgot to complete one or more of the pre- or post-questionnaires, so these data were not used in the within-subjects analysis. One participant self-disclosed being diagnosed with anxiety, and we identified their stress levels as outliers [72], so these data were removed.

Despite being prompted to take HR measurements at specific times during the study, six participants forgot or did not save their

HR measurements, so we did not include their HR data. Before analysing HR, we filtered it to ensure there were no data above 200 beats per minute (BPM) or below 25 BPM [9] (although none were found). We used Python Fitparse [11] to extract and obtain the HR data from the Garmin wrist device, and Python Pandas [69] to calculate users' incremental HR.

## 5.6 Statistical analysis and results

To incorporate confounding factors/covariates in our statistical model, and help protect against Type I errors, we employed MANCOVA and ANCOVA linear regression analyses using Python Statsmodels packages [75]. We utilized Statsmodels' Anderson-Darling, variance inflation factor, and Durbin-Watson tests [75], and examined histogram/scatterplot visualizations to ensure the MANCOVA/ANCOVA assumptions (*e.g.* linear relationships between variables, homogeneity of regressions, absence of multicollinearity, *etc.* as described in [21, 26]) were satisfied. Unless otherwise noted, each set of measures satisfied the required assumptions. We present the full output of the MANCOVA/ANCOVA analyses in the supplementary materials, and summarize the findings in the following sections.

In all of our analyses, we used an alpha level of $p<.05$. When the ANCOVA/MANCOVA analyses identified variables of interest, to determine significance, we first used Python SciPy to test the data for normality [13, 14, 90], and then performed the post-hoc tests, Mann-Whitney U [89] and independent t-test [88], as applicable.

*5.6.1 Engagement.* Through visualising the engagement data, as in Figure 9, we noticed engagement seemed higher in the audio/haptic condition. To test this, we first performed an ANCOVA. Our model included engagement as the dependent variable (DV), app type (audio/haptic or visual) as the independent variable (IV) and the busyness score and previous scanning experience (via number of previous times participants had scanned) as covariates. The model was significant ($F_{3,46}$=2.90, $R^2$=.159, p=.045), and the relationship between app type and engagement was significant (t(46)=-2.08, p=.043). There was no evidence that the covariates had significant effects on engagement.

Because app type was a significant predictor of engagement, we performed post-hoc tests on engagement and its contributors, focused attention, reward factor, and perceived usability [73]. We found **engagement** (audio/haptic,visual: $\bar{x}$=3.76,3.40; t(49)=2.24; p=.015), **reward** ($\bar{x}$=3.95,3.60; t(49)=1.78; p=.041) and **usability** ($\bar{x}$=3.95,3.57; U=396; p=.049) **were all significantly higher in the audio/haptic condition**, as shown in Figure 9 and 10. The difference for focused attention was not statistically significant.

*5.6.2 Safety proxies.* Because we did not want to put participants in danger, we used proxy variables instead of directly measuring safety. These included:

- the number of times users collided with obstacles,
- incremental HR as a proxy for stress [8, 49, 51],
- emotional affect via the circumplex model to determine whether users felt more stressed after scanning [60, 87], and
- social-evaluative, cognitive, and physical stressors via Likert scales [28, 87].

**None of the analyses indicated either condition (audio/haptic or visual) was less safe or more stressful than the other.** The supplementary materials contain detailed explanations of how we reached these conclusions—namely, similar before/after distributions and no evidence that app type predicted these variables.

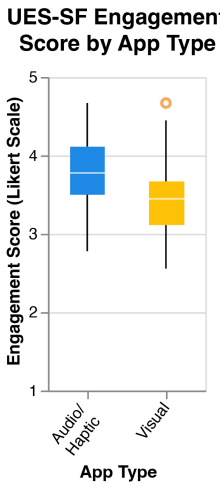


**Figure 9: The UES-SF combined engagement scores on a 5-point Likert scale. Engagement with the audio/haptic app is significantly higher than with the visual app. Note that all box-plots in this paper are quartile-based (Tukey) plots.**

We also presented participants with a list of potential stressors (which we identified in the pilot study, and are shown in the supplementary materials), and asked them which ones affected them during the scan. As shown in Figure 11, participants in the visual condition mentioned navigating crowds, not providing others with enough privacy, and feeling like the activity was not socially acceptable as stressors more often than audio/haptic users did. Those in the audio/haptic condition mentioned that the potential for thieves and physically holding the phone more frequently as stressors than visual users.

*5.6.3 User scan accuracy and scan length.* To identify the accuracy at which users followed the scanning guidance, we performed a MANCOVA analysis of the following DVs: (1) percentage of time in which the *Object Transform* was in view, (2) percentage of time the user was within the correct distance, and (3) number of times users received speed warnings for moving too quickly. We also included the time spent scanning (*i.e.* scan length) as a DV, as it is related to the speed of the scan. We analysed these with respect to the IV, app type, and the covariates: (1) number of times participants had previously scanned, (2) whether the participant encountered drift, and (3) busyness score at the time of scanning.

From the MANCOVA ($F_{1,42}$=71.2, Pillai'sTrace=.629, p<.000), there was evidence for app type having a significant effect on the DVs ($F_{1,42}$=8.07, Pillai'sTrace=.161, p=.007). After post-hoc analyses using the Mann-Whitney U test (due to each variable not satisfying normality) we found no evidence for significant differences between app type for the first three variables; however, we did find evidence for a significant difference in scan length. **Users engaged with audio/haptic scanning significantly longer than with**

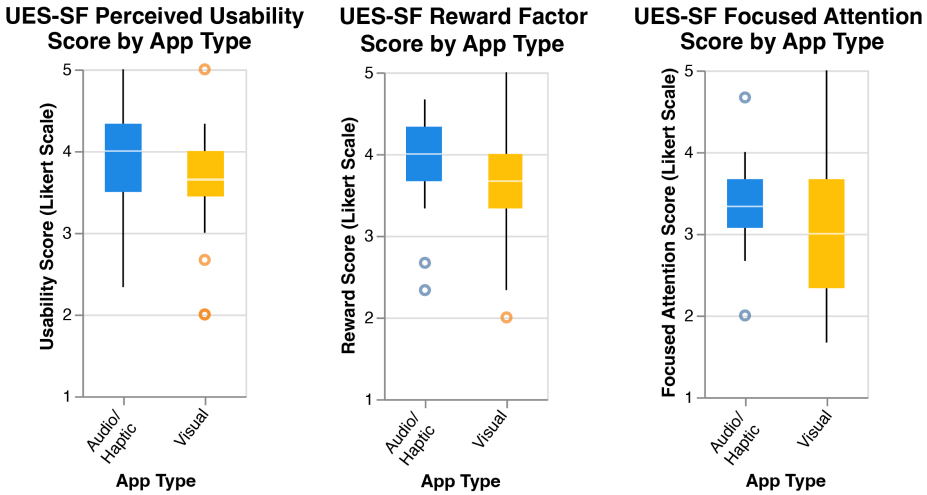


**Figure 10: The UES-SF usability, reward, and attention scores. Usability and reward with the audio/haptic app are significantly higher than with the visual app.**

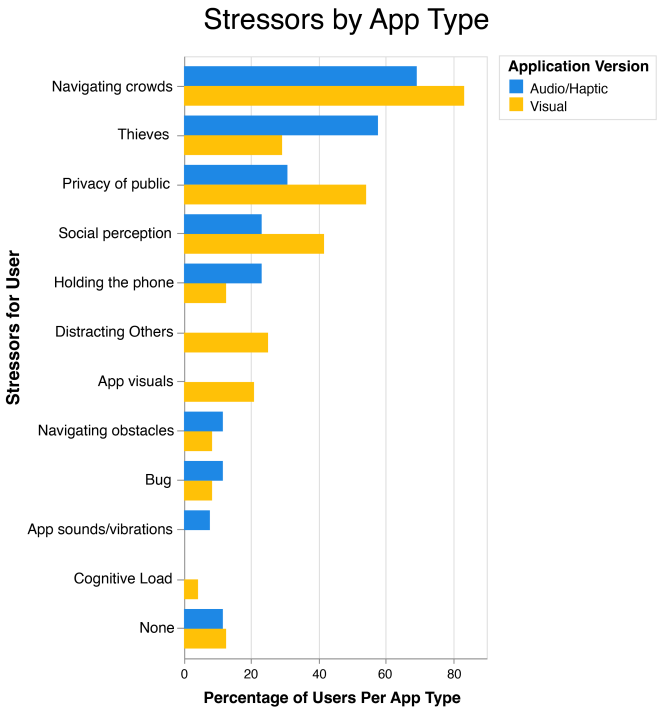


**Figure 11: The percentage of users by app type who identified particular stressors in their scanning experience. Notice how navigating crowds, others' privacy, and others' social perceptions were mentioned more frequently by visual app users than audio/haptic users.**

**Table 5: The number of participants who collided with objects or people.**

| App Type | Number Participants | Collided Once | Collided Multiple Times | Collided (at all) |
|---|---|---|---|---|
| Audio/haptic | 26 | 5 (20.83%) | 1 (4.17%) | 6 (25.00%) |
| Visual | 24 | 3 (12.50%) | 2 (8.33%) | 5 (20.83%) |

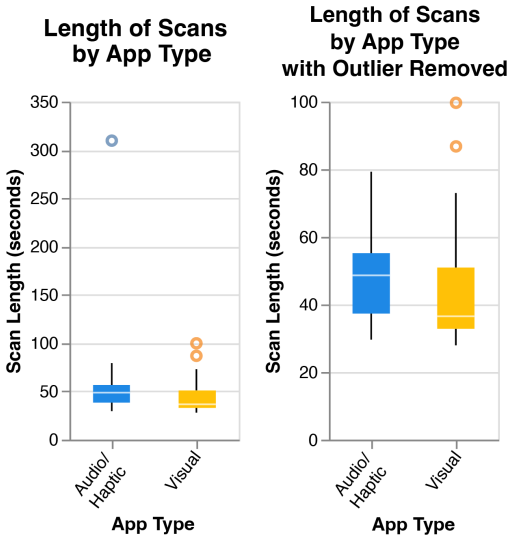


**Figure 12: Scan lengths, with and without an outlier. Users scanned for significantly longer with the audio/haptic app.**

**visual scanning** (audio/haptic,visual(seconds): $\bar{x}$=59.2,45.0; U=415; p=.02), as shown in Figure 12.

*5.6.4 Drift.* In the final questionnaire, we asked participants whether they encountered drift. As shown in Figure 13 and Table 6, **only 18% of users experienced drift**, which was a huge improvement over the pilot study, in which 83% of users experienced drift. This illustrates the effectiveness of our Landmark Tracker drift-reducing algorithm.

If participants had experienced drift, we asked them whether they were able to identify the location where the *Object Transform* had drifted using the audio/haptic or visual guidance on the app. As a side note, to illustrate the user experience of drift in the audio/haptic case, the user would be aiming the phone correctly at the physical statue; however, they would be receiving haptic feedback as if they were aiming *incorrectly*. To relocate the statue, they would find the angle at which there was no haptic feedback. The visual case is simpler: users would observe visually that the "blue track" around the *Object Transform* had moved off of the physical statue. Despite the potential for the audio/haptic condition to be more difficult, every participant who experienced drift was able to relocate the *Object transform*. This indicates that the audio/haptic guidance is robust enough for users to follow an object—not just to keep users pointing at a stationary object.

*5.6.5 User experience: Unprompted comments.* Surprisingly, around 30% of users provided unprompted, anonymous comments on the surveys. Many of the audio/haptic-condition comments were positive, and sentiment-wise, the audio/haptic-condition comments came out ahead of the visual-condition ones. One of the more thorough audio/haptic-condition comments said "having the screen off and not displaying a live image help[ed] [them] feel less concerned that others perceived [them] as invading their privacy", and that they felt "much more confident scanning while the screen remain[ed] dark" because they could tell they were "still producing a good scan using audio cues". See the supplementary materials for all comments.

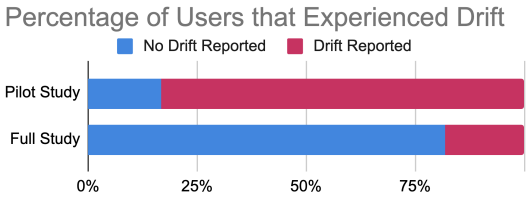


**Figure 13: The percent of participants who experienced drift in both studies. Notice the large decrease after implementing the Landmark Tracker drift-reducing algorithm in the full-study.**

**Table 6: The number of participants who experienced drift by app type. We expected drift to be similar between apps, as they used the same drift-reducing algorithm. Notice that all participants who experienced drift were able to relocate the drifted object through the provided guidance (including audio/haptic).**

| App Type | Participants | Experienced Drift | Found Object (Given Drift) |
|---|---|---|---|
| Audio/haptic | 26 | 5 (19.23%) | 5 (100%) |
| Visual | 24 | 4 (16.67%) | 4 (100%) |

## 5.7 Scanned mesh accuracy

We numerically compared scans from the user study. First we obtained a high quality reconstruction of the Agatha Christie statue, denoted as Ground Truth (GT). We then aligned each scan to this GT reconstruction and generated a reconstruction using images in each scan. Finally, we compare output meshes from the user study to the GT reconstruction and output quantitative scores. Please see the supplementary materials for the processing needed to reconstruct the ground truth, align the scans, and reconstruct the fused meshes.

*Mesh comparison.* We compare every scan's mesh with the GT to compute a Chamfer Distance and an F-score following [46]. We report average scores for every mesh and for each set of audio/haptic and visual meshes. We report these scores in Table 7. Generally, audio/haptic-condition reconstructions are better than visual-condition reconstructions on every metric.

To verify that the audio/haptic mesh accuracy was significantly better statistically, we analyzed mesh accuracy metrics using ANCOVA and a post-hoc t-test. Since ANCOVA uses linear modeling, and F-score is a nonlinear metric involving a threshold and a harmonic mean, we analyzed Chamfer Distance. The ANCOVA was significant ($F_{3,46}$=3.19, $R^2$=.172, p=.032), and the relationship between app type and Chamfer Distance was significant (t(46)=2.07, p=.044). Furthermore, the audio/haptic-condition Chamfer Distances were significantly more accurate than the visual-condition ones (audio/haptic,visual: $\bar{x}$=.055,.067; t(49)=-2.08, p=.021).

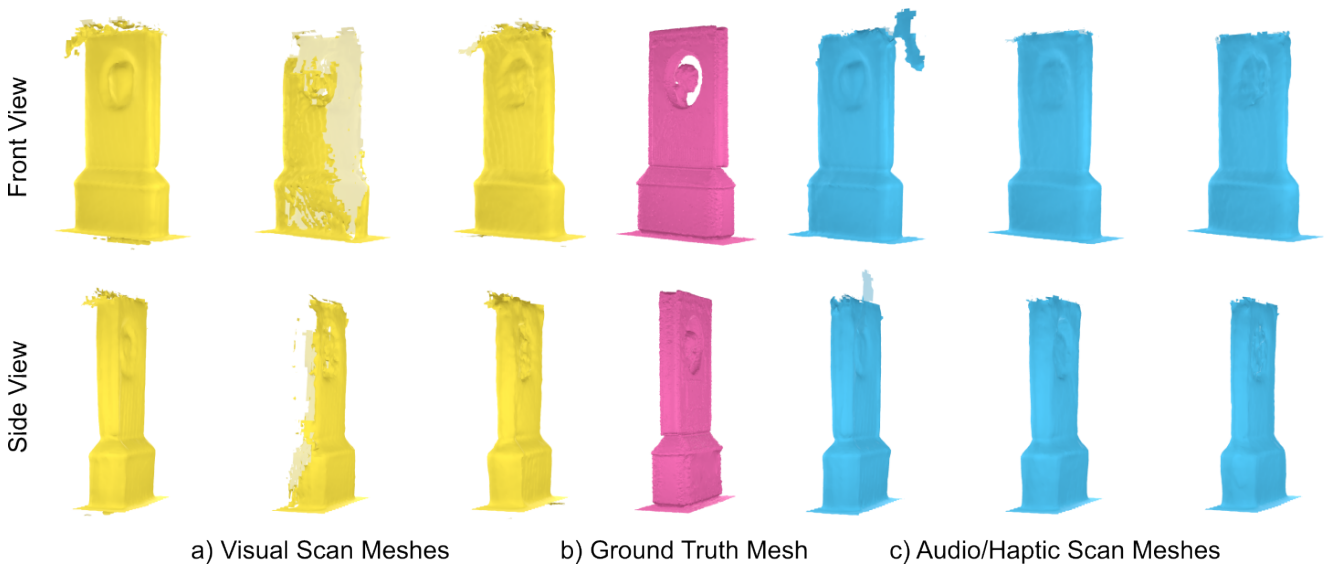


**Figure 14: b) Our high quality mesh of the statue that we use as a ground truth when evaluating reconstructions from users' scans. a) and c) are the top three scoring mesh reconstructions from both visual scans and audio/haptic scans respectively rendered from both front view and side views.**

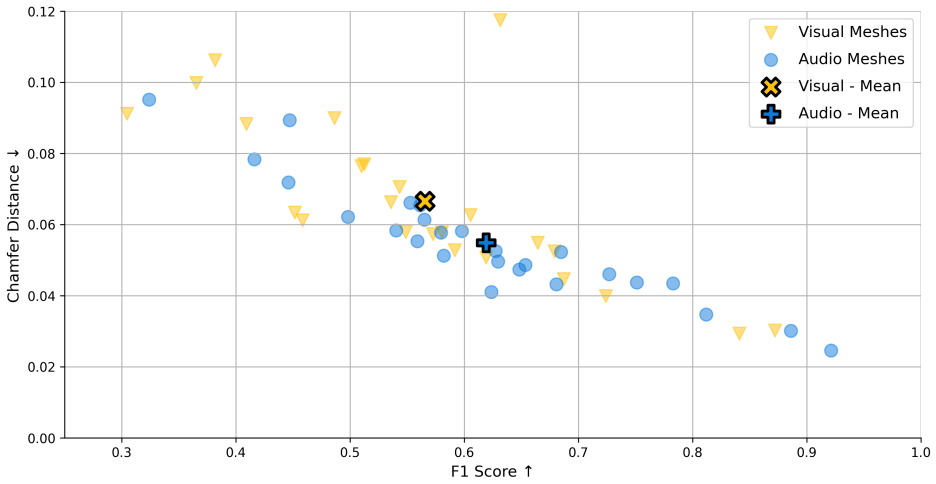


**Figure 15: We plot F-scores and Chamfer Distances for all meshes from the final study. Audio/haptic meshes have a higher (better) mean F-Score and a smaller (better) mean Chamfer Distance compared to visual meshes. Most visual meshes lie in the top left portion of the graph with a lower F-Score and a larger Chamfer Distance where four of the worst five meshes come from Visual scans.**

**Table 7: Quantitative comparison of user study meshes. Chamfer Distance is the distance between every point in one mesh and its closest neighbor in the other point cloud and vice versa—the lower the better. F-score describes the ratio of points whose distance to their closest neighbor in the other point cloud is less than a 5cm threshold—the higher the better. Please refer to [46] and the Supplementary Materials for an accurate description of what these metrics mean.**

| **Split/App Type** | Accuracy↓ (cm) | Completeness↓ (cm) | **Chamfer↓ (cm)** | Precision↑ | Recall↑ | **F-score↑** |
|---|---|---|---|---|---|---|
| All | 5.83 | 6.27 | 6.04 | 60.0% | 60.2% | 59.3% |
| Visual | 5.85 | 7.47 | 6.66 | 60.0% | 55.3% | 56.6% |
| Audio-Haptic | 5.82 | 5.17 | 5.49 | 60.1% | 64.6% | 61.9% |

# 6 DISCUSSION

Thousands of people are making 3D scans of objects, including landmarks, anatomical models, and historical artifacts, to name a few [19, 35, 47, 59, 82]. Yet, learning to scan well can be difficult [1, 71]. It's possible that existing instructions and apps are sufficient for enthusiasts and professionals, so we sought to study and improve the scanning experience for everyone else—at least for the illustrative outdoor statue setting.

While validating our hypothesis, we learned three insights related to it:

(1) Guided is better than unguided scanning. Even when users know what to do, they lose track of distance-to-POI and get distracted with "heat-of-battle" stressors. While onboarding the user to a UI with more elements is the cost of guidance, the investment pays off quickly with better scans. Guidance also increased users' feelings of confidence when scanning.
(2) Semantics-aware computer vision helps the app see the user through from start to finish. Without segmentation, it was frustrating for users to painstakingly initialize a placeholder object manually, only to have it drift off part-way through the scan. New lightweight CNN's can support stripped-down versions of emerging Segment Anything [45] types of models. Those, as we've shown, can work in concert with live geometry-focused SLAM methods. Now the user only needs to tap on the thing they wish to scan.
(3) The audio/haptic guidance was more effective than the comparable visual guidance baseline. This lesson is more nuanced. We observed statistically significant improvements in engagement, *i.e.* usability, reward factors, and attention. Further, the resulting 3D scans were surprisingly better overall. Yet, for all its benefits, audio/haptic guidance failed to reduce users' stress levels.

We also found that users of the audio/haptic app spent a little longer per scan than visual users, which may have been because this app was more engaging. This likely contributed to the improved accuracy of audio/haptic scans.

Working through the stages (see Table 1) of conceptual design [38] and the pilot showed us why the commercial scanning apps (*e.g.* [35, 66, 67]) have converged to similar interfaces. They highlighted for us, that ingenuity around the UI would only be meaningful if we could overcome the algorithmic challenges around initialization and drift. It turned out to be a prerequisite that we build the two new (at least for this context) algorithms for (i) interactive segmentation with 3D bounding-box estimation and (ii) the drift-reducing Landmark Tracker. Only 18% of users in the final study encountered any drift, compared to 83% of users in the pilot, where it was especially damaging to the scans. We strongly encourage scanning-guidance interfaces in the future to start from this substrate, or work to improve it further.

## 6.1 Limitations and future work

While we celebrate successful validation of *most* of the hypothesis, the lack of impact on stress is noteworthy. There was no significant overall change, despite the audio/haptic condition being less familiar. Hiding the video feed seemed to increase users' feelings of awareness and reduce social-evaluative stress. We speculate that like driving while listening to someone [41], scanning with either UI is an inherently demanding task requiring multi-tasking. Careful future work could further explore the safety implications of scanning. Sadly, self-reports of collisions were similar between our audio/haptic and visual conditions (see Table 5). Overall, 22% of participants collided with at least one object or person.

In this study, we chose to not include a combined audio/haptic/visual condition, as we speculated that having too many sensory modalities could increase users' cognitive load [6, 27, 68], rather than decrease stress. Nonetheless, it would be interesting to investigate the effects of scanning with all three modalities—especially on engagement and accuracy—in future work.

Although we included many potential confounding factors/covariates in our analysis and conducted a sizable (n=50) study, the unstructured, outdoor environment could have affected the results. Future work may include testing across location types, *e.g.* quieter areas *vs.* busy areas, or areas with more/fewer obstacles.

We also note that fast-paced music can increase heart rate [15]. Our relatively fast music (100 BPM) could have affected our results—and yet, we did not find that users' HRs were significantly higher in the audio/haptic condition. Future work could also include developing an app to scan alternative POIs, *e.g.* building facades, large open areas, or indoor spaces.

Evidently, there are benefits to using the audio/haptic scanning app. However, we also recognize some of these benefits (*e.g.* increased engagement) could be due to the novelty of the interface. In a longer-term study, the results may change after repeated use of the audio/haptic app. Nonetheless, we recommend the vision-assisted guidance for all scanning modes, and at least for short-term users, that they use our audio/haptic interface.

# ACKNOWLEDGMENTS

Special thanks to Diego Mazala, Charlie Houseago, Max Heimbrock, Kelly Cho, George Ash, Amy Duxbury, Jessica Nunn, Alex Morris, Summer Gu, Sen Chang, Keith Ito, Keir Rice, Thomas Hall, KP Papangelis, and Matthew Prestopino for their help with debugging, giving feedback, brainstorming, and/or organizing this project. Thanks to Vanessa Van Brummelen for her excellent, informative scanning illustration. Last but not least, thanks to all of our study participants and the reviewers for their valuable time and feedback.

# REFERENCES

[1] Daniel Andersen, Peter Villano, and Voicu Popescu. 2019. AR HMD Guidance for Controlled Hand-Held 3D Acquisition. *IEEE Transactions on Visualization and Computer Graphics* 25, 11 (2019), 3073–3082. https://doi.org/10.1109/TVCG.2019.2932172
[2] Inc. Apple. 2023. Apple Developer Documentation. https://developer.apple.com/documentation/arkit/arsession/
[3] Manuel Baer, Thomas Tregel, Samuli Laato, and Heinrich Söbke. 2022. Virtually (Re)Constructed Reality: The Representation of Physical Space in Commercial Location-Based Games. In *Proceedings of the 25th International Academic Mindtrek Conference* (Tampere, Finland) *(Academic Mindtrek '22)*. Association for Computing Machinery, New York, NY, USA, 9–22. https://doi.org/10.1145/3569219.3569339
[4] Valentin Bazarevsky and Ben Hahn. 2023. Interactive image segmentation task guide. https://developers.google.com/mediapipe/solutions/vision/interactive_segmenter
[5] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. 2019. Thematic Analysis. In *Handbook of research methods in health social sciences*, Pranee Liamputtong (Ed.). Springer, Singapore, Singapore.
[6] Roland Brünken, Jan L. Plass, and Detlev Leutner. 2004. Assessment of Cognitive Load in Multimedia Learning with Dual-Task Methodology: Auditory Load

and Modality Effects. *Instructional Science* 32, 1 (January 1 2004), 115–132. https://doi.org/10.1023/B:TRUC.0000021812.96911.c5

[7] Gary Burnett. 2000. 'Turn right at the traffic lights': The requirement for landmarks in vehicle navigation systems. *The journal of Navigation* 53, 3 (2000), 499–510.

[8] Ching Kit Chen, Barbara Cifra, Gareth J. Morgan, Taisto Sarkola, Cameron Slorach, Hui Wei, Timothy J. Bradley, Cedric Manlhiot, Brian W. McCrindle, Andrew N. Redington, Lee N. Benson, and Luc Mertens. 2016. Left Ventricular Myocardial and Hemodynamic Response to Exercise in Young Patients after Endovascular Stenting for Aortic Coarctation. *Journal of the American Society of Echocardiography* 29, 3 (2016), 237–246. https://doi.org/10.1016/j.echo.2015.11.017

[9] Caroline Christian, Elizabeth Cash, Dan A. Cohen, Christopher M. Trombley, and Cheri A. Levinson. 2023. Electrodermal Activity and Heart Rate Variability During Exposure Fear Scripts Predict Trait-Level and Momentary Social Anxiety and Eating-Disorder Symptoms in an Analogue Sample. *Clinical Psychological Science* 11, 1 (2023), 134–148. https://doi.org/10.1177/21677026221083284 arXiv:https://doi.org/10.1177/21677026221083284

[10] SeungA Chung, Kyungyeon Lee, and Uran Oh. 2021. Understanding the Two-Step Nonvisual Omnidirectional Guidance for Target Acquisition in 3D Spaces. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 339–346. https://doi.org/10.1109/ISMAR52148.2021.00050

[11] David Cooper. 2013. python-firparse Documentation. http://dtcooper.github.io/python-fitparse/

[12] Gloria Cosoli, Luca Antognoli, Valentina Veroli, and Lorenzo Scalise. 2022. Accuracy and Precision of Wearable Devices for Real-Time Monitoring of Swimming Athletes. *Sensors* 22, 13 (2022), 16 pages. https://doi.org/10.3390/s22134726

[13] Ralph B. D'Agostino. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 2 (08 1971), 341–348. https://doi.org/10.1093/biomet/58.2.341 arXiv:https://academic.oup.com/biomet/article-pdf/58/2/341/732701/58-2-341.pdf

[14] Ralph B. D'Agostino and E. S. Pearson. 1973. Tests for departure from normality. Empirical results for the distributions of b2 and b1. *Biometrika* 60, 3 (12 1973), 613–622. https://doi.org/10.1093/biomet/60.3.613 arXiv:https://academic.oup.com/biomet/article-pdf/60/3/613/576953/60-3-613.pdf

[15] C. Darki, J. Riley, D. P. Dadabhoy, A. Darki, and J. Garetto. 2022. The Effect of Classical Music on Heart Rate, Blood Pressure, and Mood. *Cureus* 14, 7 (Jul 2022), e27348.

[16] Abe Davis, Marc Levoy, and Frédo Durand. 2012. Unstructured Light Fields. *Computer Graphics Forum* 31 (2012).

[17] Davison. 2003. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings Ninth IEEE International Conference on Computer Vision*. IEEE, 1403–1410.

[18] Luca Di Angelo, Paolo Di Stefano, and Emanuele Guardiani. 2022. A review of computer-based methods for classification and reconstruction of 3D high-density scanned archaeological pottery. *Journal of Cultural Heritage* 56 (2022), 10–24. https://doi.org/10.1016/j.culher.2022.05.001

[19] Ishan Dixit, Samantha Kennedy, Joshua Piemontesi, Bruce Kennedy, and Claudia Krebs. 2019. *Which Tool Is Best: 3D Scanning or Photogrammetry – It Depends on the Task*. Springer International Publishing, Cham, 107–119. https://doi.org/10.1007/978-3-030-06070-1_9

[20] Matthew Ernst and Audrey Girouard. 2016. Exploring Haptics for Learning Bend Gestures for the Blind. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2097–2104. https://doi.org/10.1145/2851581.2892382

[21] Andy Field. 2013. *Discovering statistics using IBM SPSS statistics*. sage.

[22] Transport for London (TFL). 2023. Transport for London Crowding API. https://api-portal.tfl.gov.uk/api-details#api=crowding&operation=dayofweek

[23] Daniel Fuller, Emily Colwell, Jonathan Low, Kassia Orychock, Melissa Ann Tobin, Bo Simango, Richard Buote, Desiree Van Heerden, Hui Luan, Kimberley Cullen, Logan Slade, and Nathan G A Taylor. 2020. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR Mhealth Uhealth* 8, 9 (8 Sep 2020), e18694. https://doi.org/10.2196/18694

[24] Yasutaka Furukawa and Carlos Hernández. 2015. Multi-View Stereo: A Tutorial. *Found. Trends. Comput. Graph. Vis.* 9, 1–2 (jun 2015), 1–148.

[25] Danilo Gallo, Shreepriya Shreepriya, and Jutta Willamowski. 2020. RunAhead: Exploring Head Scanning Based Navigation for Runners. 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/10.1145/3313831.3376828

[26] G David Garson. 2012. Testing statistical assumptions.

[27] Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies* 85 (2016), 4–15. https://doi.org/10.1016/j.ijhcs.2015.08.004 Data Sonification and Sound Design in Interactive Systems.

[28] Martin Gjoreski, Mitja Luštrek, Matjaž Gams, and Hristijan Gjoreski. 2017. Monitoring stress with a wrist device using context. *Journal of Biomedical Informatics* 73 (2017), 159–170. https://doi.org/10.1016/j.jbi.2017.08.006

[29] GN Group. 2023. Jabra Elite 5. https://www.jabra.co.uk/bluetooth-headsets/jabra-elite-5#100-99181701-98

[30] Javier Hernandez, Rob R. Morris, and Rosalind W. Picard. 2011. Call Center Stress Recognition with Person-Specific Models. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 125–134.

[31] Hexagon. 2023. Immersal App - your personal metaverse. https://immersal.com/app

[32] Koki Ibukuro. 2020. Unity Taptic Plugin. https://github.com/asus4/unity-taptic-plugin

[33] Koki Ibukuro. 2023. TensorFlow Lite for Unity Samples. https://github.com/asus4/tf-lite-unity-sample

[34] Apple Inc. 2023. Meet Object Capture for iOS. https://developer.apple.com/videos/play/wwdc2023/10191/

[35] Snap Inc. 2022. Bringing Locations to Life with AR - Snap!'s Custom Landmarker Creator. https://eng.snap.com/life-with-ar-landmarkers

[36] Ananya Ipsita, Hao Li, Runlin Duan, Yuanzhi Cao, Subramanian Chidambaram, Min Liu, and Karthik Ramani. 2021. VRFromX: from scanned reality to interactive virtual experience with human-in-the-loop. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[37] Jérôme Isabelle and Denis Laurendeau. 2021. A Mixed Reality Interface for Handheld 3D Scanners. In *Human Interaction, Emerging Technologies and Future Applications III*, Tareq Ahram, Redha Taiar, Karine Langlois, and Arnaud Choplin (Eds.). Springer International Publishing, Cham, 189–194.

[38] Daniel Jackson. 2015. Towards a Theory of Conceptual Design for Software. In *2015 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software (Onward!)* (Pittsburgh, PA, USA) *(Onward! 2015)*. Association for Computing Machinery, New York, NY, USA, 282–296. https://doi.org/10.1145/2814228.2814248

[39] Daniel Jackson. 2021. *The Essence of Software: Why Concepts Matter for Great Design*. Princeton University Press, Princeton. https://doi.org/10.1515/9780691230542

[40] Matt Jones, Steve Jones, Gareth Bradley, Nigel Warren, David Bainbridge, and Geoff Holmes. 2008. ONTRACK: Dynamically Adapting Music Playback to Support Navigation. *Personal Ubiquitous Comput.* 12, 7 (oct 2008), 513–525. https://doi.org/10.1007/s00779-007-0155-2

[41] Marcel Adam Just, T. Anderson Keller, and Jacquelyn Cynkar. 2008. A decrease in brain activation associated with driving when listening to someone speak. *Brain Research* 1205 (2008), 70–80.

[42] Akankasha Kathuria. 2023. Analysis Of 3d Scanning And Reconstruction Techniques For Realistic Object Animation. *Elementary Education Online* 20, 3 (2023), 4655–4655.

[43] T. P. Kersten, M. Lindstaedt, and D. Starosta. 2018. COMPARATIVE GEOMETRICAL ACCURACY INVESTIGATIONS OF HAND-HELD 3D SCANNING SYSTEMS – AN UPDATE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2 (2018), 487–494. https://doi.org/10.5194/isprs-archives-XLII-2-487-2018

[44] Minju Kim and Jungjin Lee. 2019. PicMe: Interactive Visual Guidance for Taking Requested Photo Composition. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019). https://api.semanticscholar.org/CorpusID:140302928

[45] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2023).

[46] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM Transactions on Graphics* 36, 4 (2017).

[47] Jarrod Knibbe, Kenton P. O'Hara, Angeliki Chrysanthi, Mark T. Marshall, Peter D. Bennett, Graeme Earl, Shahram Izadi, and Mike Fraser. 2014. Quick and Dirty: Streamlined 3D Scanning in Archaeology. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1366–1376. https://doi.org/10.1145/2531602.2531669

[48] Samuli Laato and Thomas Tregel. 2023. Into the Unown: Improving location-based gamified crowdsourcing solutions for geo data gathering. *Entertainment Computing* 46 (2023), 100575. https://doi.org/10.1016/j.entcom.2023.100575

[49] Taija MM Lahtinen, Jukka P Koskelo, Tomi Laitinen, and Tuomo K Leino. 2007. Heart rate and performance during combat missions in a flight simulator. *Aviation, space, and environmental medicine* 78, 4 (2007), 387–391.

[50] Fabian Langguth and Michael Goesele. 2013. Guided Capturing of Multi-view Stereo Datasets. In *Eurographics 2013 - Short Papers*. The Eurographics Association.

[51] Yung-Hui Lee and Bor-Shong Liu. 2003. Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine* 74, 10 (2003), 1078–1084.

[52] John J Leonard and Hugh F Durrant-Whyte. 1991. Simultaneous map building and localization for an autonomous mobile robot.. In *IROS*, Vol. 3. 1442–1447.

[53] Niantic Lightship. 2023. Building ARDK Apps for IOS. https://lightship.dev/docs/archive/ardk/ardk_fundamentals/building_ios.html#steps

[54] Garmin Ltd. 2023. Garmin Venu 2 Plus: Health and Fitness Smartwatch with GPS. https://www.garmin.com/en-GB/p/730659

[55] Hanuma Teja Maddali and Amanda Lazar. 2023. Understanding Context to Capture When Reconstructing Meaningful Spaces for Remote Instruction and Connecting in XR. 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 275, 18 pages. https://doi.org/10.1145/3544548.3581243

[56] Georgios N. Marentakis and Stephen A. Brewster. 2005. Effects of Reproduction Equipment on Interaction with a Spatial Audio Interface. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1625–1628. https://doi.org/10.1145/1056808.1056982

[57] Emanuele Marino, Fabio Bruno, Loris Barbieri, and Antonio Lagudi. 2022. Benchmarking Built-In Tracking Systems for Indoor AR Applications on Popular Mobile Devices. *Sensors* 22, 14 (2022). https://doi.org/10.3390/s22145382

[58] Bob-Antoine J. Menelas, Lorenzo Picinali, Patrick Bourdot, and Brian F. G. Katz. [n. d.]. Non-visual identification, localization, and selection of entities of interest in a 3D environment. *Journal on Multimodal User Interfaces* 8, 3 ([n. d.]). https://doi.org/10.1007/s12193-014-0148-1

[59] Peter Mooney and Padraig Corcoran. 2014. Has OpenStreetMap a role in Digital Earth applications? *International Journal of Digital Earth* 7, 7 (2014), 534–553.

[60] Jon D Morris. 1995. Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. *Journal of advertising research* 35, 6 (1995), 63–68.

[61] Judith Mwakalonge, Saidi Siuhi, and Jamario White. 2015. Distracted walking: Examining the extent to pedestrian safety problems. *Journal of Traffic and Transportation Engineering (English Edition)* 2, 5 (2015), 327–337.

[62] Michael Nebeling, Janet Nebeling, Ao Yu, and Rob Rumble. 2018. ProtoAR: Rapid Physical-Digital Prototyping of Mobile Augmented Reality Applications. 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, 1–12. https://doi.org/10.1145/3173574.3173927

[63] Inc. Niantic. 2022. How to Review Wayspots (Wayfarer Points of Interest). https://niantic.helpshift.com/hc/en/21-wayfarer/faq/2062-reviewing-a-suggested-wayspot-edit/

[64] Inc. Niantic. 2022. Ingress Portal Scanning. https://niantic.helpshift.com/hc/en/3-ingress/faq/2398-portal-scanning/?han=1

[65] Inc. Niantic. 2023. Lightship ARDK. https://lightship.dev/products/ardk

[66] Inc. Niantic. 2023. Niantic Wayfarer. https://wayfarer.nianticlabs.com/new/

[67] Inc. Niantic. 2023. Scaniverse: Capture life in 3D. https://scaniverse.com/

[68] Tomi Nukarinen, Roope Raisamo, Ahmed Farooq, Grigori Evreinov, and Veikko Surakka. 2014. Effects of Directional Haptic and Non-Speech Audio Cues in a Cognitively Demanding Navigation Task. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (Helsinki, Finland) *(NordiCHI '14)*. Association for Computing Machinery, New York, NY, USA, 61–64. https://doi.org/10.1145/2639189.2639231

[69] Inc. NumFOCUS. 2023. pandas. https://pandas.pydata.org/

[70] Stable Diffusion Online. 2023. Stable Diffusion Online. https://stablediffusionweb.com/

[71] J. Ortiz-Sanz, M. Gil-Docampo, T. Rego-Sanmartín, M. Arza-García, and G. Tucci. 2021. A PBeL for training non-experts in mobile-based photogrammetry and accurate 3-D recording of small-size/non-complex objects. *Measurement* 178 (2021), 109338. https://doi.org/10.1016/j.measurement.2021.109338

[72] Jason W Osborne and Amy Overbay. 2004. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation* 9, 1 (2004), 6.

[73] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[74] Qi Pan, Gerhard Reitmayr, and Tom W. Drummond. 2009. Interactive model reconstruction with user guidance. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Orlando, Florida, USA, 209–210. https://doi.org/10.1109/ISMAR.2009.5336460

[75] Josef Perktold, Skipper Seabold, and Jonathan Taylor. 2023. statsmodels. https://www.statsmodels.org/devel/index.html

[76] Grigori D. Pintilie and Wolfgang Stuerzlinger. 2013. An Evaluation of Interactive and Automated Next Best View Methods in 3D Scanning. *Computer-Aided Design and Applications* 10, 2 (2013), 279–291. https://doi.org/10.3722/cadaps.2013.279-291 arXiv:https://www.tandfonline.com/doi/pdf/10.3722/cadaps.2013.279-291

[77] Anita Pollak, Mateusz Paliga, Matias M. Pulopulos, Barbara Kozusznik, and Malgorzata W. Kozusznik. 2020. Stress in manual and autonomous modes of collaboration with a cobot. *Computers in Human Behavior* 112 (2020), 106469. https://doi.org/10.1016/j.chb.2020.106469

[78] 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

[79] London Remembers. 2023. Sculpture: Agatha Christie Book. https://www.londonremembers.com/memorials/agatha-christie-book

[80] Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. 2002. Real-Time 3D Model Acquisition. *ACM Trans. Graph.* 21, 3 (jul 2002), 438–446. https://doi.org/10.1145/566654.566600

[81] Sheila M. Ryan, Ary L. Goldberger, Steven M. Pincus, Joseph Mietus, and Lewis A. Lipsitz. 1994. Gender- and age-related differences in heart rate dynamics: Are women more complex than men? *Journal of the American College of Cardiology* 24, 7 (1994), 1700–1707. https://doi.org/10.1016/0735-1097(94)90177-5

[82] G.-Fivos Sargentis, Evangelia Frangedaki, Michalis Chiotinis, Demetris Koutsoyiannis, Stephanos Camarinopoulos, Alexios Camarinopoulos, and Nikos D. Lagaros. 2022. 3D Scanning/Printing: A Technological Stride in Sculpture. *Technologies* 10, 1 (2022). https://doi.org/10.3390/technologies10010009

[83] Mohamed Sayed, Robert Cinca, Enrico Costanza, and Gabriel Brostow. 2022. LookOut! Interactive Camera Gimbal Controller for Filming Long Takes. *ACM Trans. Graph.* 41, 3, Article 30 (mar 2022), 16 pages. https://doi.org/10.1145/3506693

[84] Mohamed Sayed, John Gibson, Jamie Watson, Victor Prisacariu, Michael Firman, and Clément Godard. 2022. SimpleRecon: 3D reconstruction without 3D convolutions. In *European Conference on Computer Vision*. Springer, 1–19.

[85] Morgan Klaus Scheuerman. 2020. HCI Guidelines for Gender Equity and Inclusivity. https://www.morgan-klaus.com/gender-guidelines.html

[86] Maximilian Schirmer, Johannes Hartmann, Sven Bertel, and Florian Echtler. 2015. Shoe Me the Way: A Shoe-Based Tactile Interface for Eyes-Free Urban Navigation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 327–336. https://doi.org/10.1145/2785830.2785832

[87] Philip Schmidt, Attila Reiss, Robert Dürichen, and Kristof Van Laerhoven. 2019. Wearable-Based Affect Recognition—A Review. *Sensors* 19, 19 (2019), 42 pages. https://doi.org/10.3390/s19194079

[88] SciPy. 2023. Independent t-test. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

[89] SciPy. 2023. Mann-Whitney U. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html

[90] SciPy. 2023. Scipy Normal Test. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html#r7bf2e556f491-2

[91] Marcos Seefelder and Daniel Duckworth. 2023. *Reconstructing indoor spaces with NeRF – Google Research Blog*. https://blog.research.google/2023/06/reconstructing-indoor-spaces-with-nerf.html

[92] Wang Shunli, Hu Qingwu, Wang Shaohua, Zhao Pengcheng, and A.I. Mingyao. 2018. A 3D Reconstruction and Visualization App Using Monocular Vision Service. In *2018 26th International Conference on Geoinformatics*. 1–5. https://doi.org/10.1109/GEOINFORMATICS.2018.8557103

[93] Keng Hua Sing and Wei Xie. 2016. Garden: A Mixed Reality Experience Combining Virtual Reality and 3D Reconstruction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 180–183. https://doi.org/10.1145/2851581.2890370

[94] Petros Skapinakis. 2014. *Spielberger State-Trait Anxiety Inventory*. Springer Netherlands, Dordrecht, 6261–6264. https://doi.org/10.1007/978-94-007-0753-5_2825

[95] A. Somogyi, A. Barsi, B. Molnar, and T. Lovas. 2016. CROWDSOURCING BASED 3D MODELING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B5 (2016), 587–590. https://doi.org/10.5194/isprs-archives-XLI-B5-587-2016

[96] Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, and Hujun Bao. 2021. NeuralRecon: Real-time coherent 3D reconstruction from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15598–15607.

[97] Unity Technologies. 2023. Unity. https://unity.com/

[98] Sebastian Thrun. 2008. *Simultaneous Localization and Mapping*. Springer Berlin Heidelberg, Berlin, Heidelberg, 13–41. https://doi.org/10.1007/978-3-540-75388-9_3

[99] timouse. 2022. Freesound: "melodic beat 220619.wav". https://freesound.org/people/timouse/sounds/640097/

[100] Jessica Van Brummelen, Marie O'Brien, Dominique Gruyer, and Homayoun Najjaran. 2018. Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies* 89 (2018),

384–406. https://doi.org/10.1016/j.trc.2018.02.012

[101] Vanessa Van Brummelen. 2023. Guidance for 3D Scanning of Landmarks Illustration.

[102] Yolanda Vazquez-Alvarez and Stephen Brewster. 2009. Investigating Background & Foreground Interactions Using Spatial Audio Cues. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 3823–3828. https://doi.org/10.1145/1520340.1520578

[103] Styliani Verykokou and Charalabos Ioannidis. 2023. An Overview on Image-Based and Scanner-Based 3D Modeling Technologies. *Sensors* 23 (01 2023), 596. https://doi.org/10.3390/s23020596

[104] Jan B. Vornhagen, April Tyack, and Elisa D. Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 4–18. https://doi.org/10.1145/3410404.3414229

[105] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2021. DistanciAR: Authoring Site-Specific Augmented Reality Experiences for Remote Environments. 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 411, 12 pages. https://doi.org/10.1145/3411764.3445552

[106] Gregory F Welch. 2020. Kalman filter. *Computer Vision: A Reference Guide* (2020), 1–3.

[107] Wikipedia. 2023. Agatha Christie Memorial. https://en.wikipedia.org/wiki/Agatha_Christie_Memorial

[108] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. https://doi.org/10.1145/2907069

[109] Yi-Chin Wu, Liwei Chan, and Wen-Chieh Lin. 2019. Tangible and Visible 3D Object Reconstruction in Augmented Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Beijing, China, 26–36. https://doi.org/10.1109/ISMAR.2019.00-30

[110] Zhongyuan Yu, Daniel Zeidler, Victor Victor, and Matthew Mcginity. 2023. Dynascape: Immersive Authoring of Real-World Dynamic Scenes with Spatially Tracked RGB-D Videos. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology* (<conf-loc>, <city>Christchurch</city>, <country>New Zealand</country>, </conf-loc>) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. https://doi.org/10.1145/3611659.3615718

[111] Yanan Zhang, R. Glenn Weaver, Bridget Armstrong, Sarah Burkart, Shuxin Zhang, and Michael W. Beets. 2020. Validity of Wrist-Worn photoplethysmography devices to measure heart rate: A systematic review and meta-analysis. *Journal of Sports Sciences* 38, 17 (2020), 2021–2034. https://doi.org/10.1080/02640414.2020.1767348 arXiv:https://doi.org/10.1080/02640414.2020.1767348 PMID: 32552580.

[112] Haojie Zhao, Junsong Chen, Lijun Wang, and Huchuan Lu. 2023. ARKitTrack: A New Diverse Dataset for Tracking Using Mobile RGB-D Data. arXiv:2303.13885 [cs.CV]

[113] Zhengjuan Zhou. 2015. HeadsUp: Keeping Pedestrian Phone Addicts from Dangers Using Mobile Phone Sensors. *Int. J. Distrib. Sen. Netw.*, Article 5 (jan 2015), 1 pages.

[114] Robert Zlot, Michael Bosse, Kelly Greenop, Zbigniew Jarzab, Emily Juckes, and Jonathan Roberts. 2014. Efficiently capturing large, complex cultural heritage sites with a handheld mobile 3D laser mapping system. *Journal of Cultural Heritage* 15, 6 (2014), 670–678. https://doi.org/10.1016/j.culher.2013.11.009