Viability of a tangible tabletop for industry storyboarding

Katherine Rix
University of Cape Town
katherine.rix@alumni.uct.ac.za

ABSTRACT

The intuitiveness and ease of learning of tangible user interfaces (TUIs) is well-established in HCI literature. Despite this, few TUI-based applications have been developed for industry, leaving open the question of their suitability to high-expertise tasks. To investigate this, we tailored a TUI-based storyboarding application to purpose through expert interview, prototyping and heuristic evaluation before presenting it to animation professionals. Twenty participants used the system to create 3D scenes from storyboard sketches. An all-positive system usability scale (SUS) was administered, benchmarking the system's usability at 78.0 or “good”. A second, diagnostic questionnaire revealed that the system was fun to use but too slow - professionals have strong time-saving requirements due to deadline pressure. Furthermore, observed user behavior and qualitative post-session interviews strongly suggest that, even with performance improvements, applicability to the animation industry is severely curtailed by insufficiently fine-grained control, lack of model flexibility, and lack of innate mechanisms for common, critical actions like “undo”.

CCS Concepts

• **Applied computing ~ Media arts** • **Hardware ~ Tactile and hand-based interfaces** • *Human-centered computing ~ Graphics input devices* • *Human-centered computing ~ Empirical studies in HCI*

Keywords

3D animation; lo-fi animation; tabletop TUI; gamepad; usability; industry suitability

1. INTRODUCTION

Previsualization (or previz) is a low-fidelity step in the pre-production phase of making an animated film. It involves setting 3D assets in motion along basic motion paths. At its lowest fidelity, previz has no articulated animation such as limb or mouth movement. Furthermore, the paths that characters follow are simplistic and representative rather than natural [Lwabona 2015]. However, this provides enough detail for experienced animators and directors to confirm timing, framing and scene layout before the production phase, when high-fidelity animation begins.

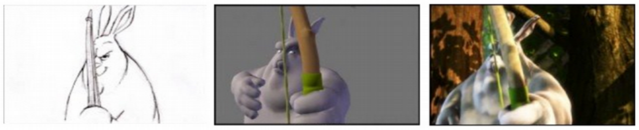


Figure 1. An example of the multi-stage animation process. From left to right: storyboard image, previz frame and final production frame in *Big Buck Bunny* [Goedegebure et al. 2008]. Note how the visible detail increases at each stage.

Previz is based on hand-drawn storyboards and is typically put together by the same 3D artists who perform high-fidelity animation tasks, using the same complex software (e.g. *Blender*[1] or *Maya*[2]). Thus, despite its lo-fi requirements, previz requires a high degree of user training.

This suggests that a less complex tool might be suitable to achieve this first transition from sketches to 3D scene data. Such a tool would lower previz training requirements in industry, and democratize [Tanenbaum et al. 2013] the creation of lo-fi animation.

Firstly, we can isolate the format transition task by removing the timing component of previsualization, reducing it to sequence of static keyframes: a storyboard. (This also removes the requirement for scene layout to remain coherent between frames.) This scene-based storyboarding activity would result in two outputs: a sequence of 2D frames (rendered using 3D assets), and the associated scene layout for each of those frames.

Secondly, we can lower training requirements and increase usability by providing users with an input/output modality that leverages their existing “reality-based” skills [Jacob et al. 2008]. Tangible user interfaces (TUIs) offer such a modality by embodying some of the input/output in physical objects that the user manipulates [Ishii 2008a, 2008b].

We propose a tabletop TUI that uses 3D-printed character models placed on a desk in front of a depth sensor, a simple GUI presented via a monitor, and an untethered gamepad for system control. The position and orientation of models is captured by the depth sensor, and the real-world scene is recreated in the GUI with the same virtual assets used to print the physical models. By rearranging the models on the tabletop, the user can create a storyboard sequence of static images. (See *Appendix A* for an comic illustrating system functionality. Note that the comic depicts a wearable input device which was later replaced by a gamepad.)

Although the theory of tangible modalities is well-supported by the literature, applications of reality-based systems are uncommon and not always successful [Norman 2010]. We therefore present not only the final prototype, but highlights from our design process, a robust usability benchmark and some discussion of industry applicability based on a qualitative evaluation.

Katherine Rix, 2015. *Viability of a tangible tabletop for industrial storyboarding*. Honours report, University of Cape Town. 13 pages. November 2015. http://people.cs.uct.ac.za/~previz2015/tangible/

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

[1] http://www.blender.org

[2] http://www.autodesk.com/products/maya/overview

2. PREVIOUS WORK

2.1 Lo-fi animation

Previous studies investigating lo-fi animation interfaces have also aimed to reduce the training requirements for system operation. These include symbolic drawing [Figueroa et al. 2014] and a touch-based digital tabletop [Lwabona 2015]. Other alternate interfaces whose usability has not been formally studied include digital sketching within the traditional high-detail interface [Roosendaal 2015a] and a proposal for a modifiable GUI, which could be reduced in complexity for entry-level users [Roosendaal 2015b, 2015c].

2.2 Tangible interfaces

Ideally, physical interfaces would be made from “radical atoms”, a hypothetical material that mimics the versatility of digital data [Ishii et al. 2012]. Practical implementations can only approximate radical atoms, typically by complementing their physical components with some kind of digital sensing system that is largely invisible to the user [Ishii 2008a, Shaer and Hornecker 2010]. For example, a computer vision system can interpret the location and orientation of physical objects. In such an approximation, versatility is traded to leverage the user's spatial and body awareness, and provide immediate haptic-visual feedback. [Ishii 2008b, Jacob et al. 2008]

2.3 Usability benchmarks

While the reality-based interaction framework [Jacob et al. 2008] and prior tangible studies provide a theoretical foundation for constructing an interface, a usability study is required to measure the effectiveness of our proposed system-task pairing (tangible-storyboarding).

The system usability scale (SUS) [Brooke 1996, 2013] is a usability instrument with broad applicability, making it suitable as an objective usability measure [Bangor et al. 2008, Sauro 2011]. Survey results [Bangor et al. 2008, Sauro 2011] allow us to compare systems on a universal scale, even when they have distinct interfaces or target distinct tasks. A raw score of 80 suggests that a system lies in the top 10% of systems surveyed [Sauro 2011], and that subjects are more likely to recommend the system to others [Sauro 2010] – this would be an ideal result for a refined system. A score can otherwise be interpreted according to an adjectival scale [Bangor et al. 2009], ranked from “worst imaginable” to “best imaginable”, which correspond to a scaled percentile ranking [Sauro 2011].

Sauro [2011] suggests that a sample size of 20 is required for a statistically significant SUS benchmark, suitable for comparison to survey results. This benchmark value can be treated as a lower bound on the *potential* usability of a system: usability will necessarily improve when interface flaws are correctly identified and fixed, and usability will also improve with repeated use (until some upper bound is reached [Sauro 2011]).

Finally, an all-positive reframing of the SUS reduces response and scoring errors without jeopardizing validity [Sauro and Lewis 2011], and so is recommended for new studies [Sauro 2011, Brooke 2013].

3. DESIGN

Our design follows the “tabletop TUI” or “workbench” approach [Ishii 2008a]. The user places physical models of the characters and scene camera on a tabletop. The system interprets their placement, and composes a storyboard frame from the perspective of the model camera. The frame is rendered on a monitor, as seen in Figure 2. The user repeats this activity for each frame in the storyboard sequence.

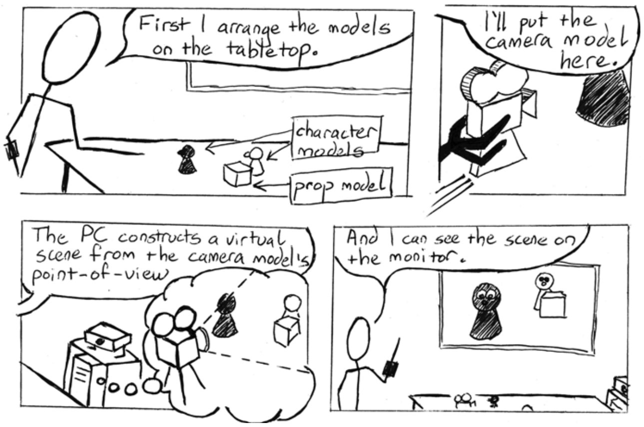


Figure 2. Excerpt from illustrative comic demonstrating system functionality (full comic in *Appendix A*).

If our interface were made of radical atoms, the models would fully embody their scene data, and direct manipulation of the models would be immediately reflected in the scene image. To approximate this ideal interface, we rely on a registration module[3], which uses a depth sensor [Shaer and Hornecker 2010] to determine the relative positions and rotations of the physical models.

Finally, our system provides the user with a non-tangible input device (a “remote control” [Ishii 2008a]) to manipulate the intangible data (the sequence of frames). This device is used to create, read, update and delete frames, and navigate along the frame sequence. Early iterations used a semaphoric[4] armband (*Myo armband*[5]), and later a gamepad (*Xbox 360 controller*).

3.1 Paper prototype and heuristic evaluation

We designed a prototype [Rettig 1994] that presents the user with paper models and a hand-drawn screen. An inactive depth sensor was placed nearby to simulate its presence in the final implementation. An inactive semaphoric armband was used as a stand-in input device.

---

[3] Registration module provided by David Seaward, with further refinements and optimizations made for this study.

[4] Semaphoric input allows users to interact with a system by making particular gestures with their hands or bodies. The use of symbolic gestures distinguishes it from other gestural modalities, such as deitic input (pointing) or direct manipulation. [O'Hara et al. 2013]

[5] https://www.myo.com

The wizard-of-oz paper system afforded the user a high degree of control over the scene, including:

- **(Almost) 6 degrees of freedom.** Models could be placed anywhere on the tabletop and rotated in any direction. Gravity was a partial limitation on Y placement (models could not be placed in mid-air, but they could be attached to other models, for example, hanging a character from the branch of a tree).
- **Sensor occlusion allowed.** Models could be placed in front of one another with respect to the depth sensor, ignoring any potential impediment.
- **Exact input interpretation.** A trained human observer interpreted the users' semaphoric input.
- **Exact scene construction.** A trained human observer constructed the storyboard frame based on the tabletop layout. The frame representation was low-fidelity, allowing the user to assume exact reconstruction. [Rettig 1994]
- **Rapid feedback.** System feedback was performed by a trained human observer, achieving a response time of less than 5 seconds.

A positive heuristic assessment from 5 participants [Nielsen and Molich 1990, Nielsen 1995] with no animation background [Molapo and Marsden 2013] confirmed that our initial design was acceptable and we proceeded to implementation.

Early prototyping of an active Myo armband for timeline navigation suggested that while its semaphoric repertoire was adequate, the device could not cope with rapid handover between participants, and so was not suitable for our experimental design. A decision was taken to switch to an Xbox 360 controller. The button mapping for this input device was developed through further heuristic evaluation.

3.2 Implementation

For our final implementation, we took character assets[6] [7] from *Big Buck Bunny* [Goedegebure et al. 2008], a high-quality 3D animation. We printed the assets and a camera prop using a 3D printer (*Ultimaker 2*[8]). We placed these printed models on a tabletop in front of an active depth sensor (*Kinect for Xbox One*[9]). A registration module interpreted the location and rotation of the models. The system placed their virtual counterparts in the same location and rotation in a 3D scene displayed on a monitor. Frame manipulation (create/read/update/delete/navigate) was presented via a simple GUI and controlled via a gamepad (*Xbox 360 controller*).

The following design principles were applied:

- **Verisimilitude of assets.** The 3D characters were taken from a high-quality, published animated short. This meant they were designed for visual aesthetics rather than directly for tangible use or printability. An industry-oriented system should not limit aesthetic decisions based on technical constraints, and prototyping with context-appropriate assets focuses feedback on user needs [Molapo and Marsden 2013].
- **Figurative assets.** We elected to print the same models used to render the scenes, rather than abstract standins. This is in line with the “input = output” paradigm of TUI systems [Ishii 2008a], specifically as applied to visual judgments while laying out a scene. Visual judgments might be made according to relative model height, hip placement, foot placement or other feature-based measures that would be lost in abstraction.
- **Appropriate handling size.** The physical size of the models (between 8 and 11cm tall) was selected based on heuristic assessment from the paper prototype phase.
- **Accessible multimodal feedback.** Reality-based systems typically do not rely on a single modality for all input/output [Oviatt 1999]. Our system provided visual and audio feedback on state and progress. Visual widgets and other visual components were larger than traditional desktop widgets, given the proposed operating distance (user viewing the screen across an office desk while standing).
- **Gamepad mapping.** Our button mapping was designed based on heuristic evaluations from participants with and without gamepad experience. (Mapping conventions have evolved in the gaming context, the original context for gamepad controllers. Our heuristic evaluation confirmed which navigation conventions were intuitive for users without gamepad experience, which were not appropriate, and which conventions had strong expectations associated with them for users with gamepad experience.)

Figure 3. Button mapping for system control and storyboard navigation. Note that users may move left and right using buttons on the back of the controller (a gaming convention) or the directional pad (a more immediate and intuitive option).

Before starting the usability survey we identified some key limitations in the as-built prototype, which we anticipated would have a negative impact on the usability and/or suitability:

- **3 degrees of freedom.** As a simplifying assumption, the registration module assumes models will be placed only on

[6] *Big Buck Bunny* character models are available to paying members at https://cloud.blender.org – fees support the Blender Foundation.

[7] *Big Buck Bunny* character models are available at no cost at http://graphicall.org/bbb/index.php (under “chars”)

[8] https://ultimaker.com

[9] http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one

the XZ plane and rotated only about their Y axis. We account for this limitation by not requiring the users to recreate scenes that the system cannot interpret.

- **Sensor occlusion not allowed.** A model cannot be placed between another model and the sensor, limiting the freedom of placement on the stage.
- **Scene construction: limited accuracy.** The registration module sometimes misidentifies models, and sometimes returns inexact rotation results. Position results seem adequate.
- **Scene construction: limited stage size.** The area in which models can be placed is limited by the accuracy of the sensor and registration module. We account for this limitation by clearly marking the "stage" area.
- **Scene construction: all models must be present.** In order to aid identification, all models have to be present on the stage. We account for this limitation by marking a convenient "out of shot" alley on the stage.
- **Delayed feedback.** The registration module initially took as long as 5 minutes to visualize the scene on the monitor. We performed optimizations to reduce this time to 15-30 seconds. The remaining delay is still a limitation, which we mitigate by providing progress feedback in a dialog box.

4. AIMS AND METHODOLOGY

Our goal is to confirm the usability of this tangible-storyboarding system, and investigate the viability of this system-task pairing in an industry context (a professional animation studio).

Participants were asked to recreate a hand-drawn storyboard sequence using the tangible prototype. We took a mixed-methods approach in assessing the usability and suitability of the system.

4.1 System-task usability

We performed an all-positive SUS benchmark to establish an objective, quantitative measure of usability. Our assumption is that a prototype system with an above-average usability score suggests a promising avenue that may lead to a system with an ideal score.

Hypothesis 1. When used for a storyboarding activity, the prototype tangible interface will achieve an above-average SUS benchmark.

A failed hypothesis would suggest that a tangible interface is not well-suited for storyboarding, or that our system needs refinement before drawing a firm conclusion.

4.2 Industry suitability

In parallel, we performed a qualitative investigation to establish the suitability of the system for everyday industry use.

In earlier iterations we had difficulty describing the system verbally to potential participants and fellow researchers. This raised the possibility that we were using unintuitive language. To probe this question, participants were presented with a rough sketch of the system in front of them. Ten components of the system were numbered. Participants were asked to label these components using their own words.

Further qualitative investigation took the form of "think-aloud" feedback while participants used the system, a diagnostic feedback form, and after-session interviews.

Hypothesis 2. Observation of and feedback from users at a professional animation studio will confirm the suitability of a tangible interface to lo-fi animation.

4.3 Pilot

We ran an initial 5 participant pilot to assess this design. Some participants were confused when trying to achieve shots that had only one of the two characters in the camera's view. In response we added an "out of shot" alley to the stage, to aid with scene organization.

5. RESULTS

5.1 Sample demographics

We performed our system benchmark with 20 participants over the course of 4 days, in a familiar environment (their place of work).

13 participants were male and 7 female. The minimum reported age was 23, the maximum 45 and the mean reported age was 31.5. Age and gender are not known to have a significant effect on SUS scores [Sauro 2011].

Participants were categorized according to occupation and industry skillset: 6 participants were classified as "2D artists", 9 as "3D artists" and 2 as "directors". The remaining 3 participants were classified as "non-artists" (those in supporting roles, or with limited or no direct animation experience). 2D artists had a digital and non-digital (sketching) skillset.

14 users reported neutral-or-better familiarity with gamepads. Of these, 8 (40%) reported better-than-neutral familiarity.

14 users reported neutral-or-better familiarity with animation software such as *Blender* and *Maya*. Of these, 12 (60%) reported better-than-neutral familiarity.

5.2 System-task viability

In a benchmark all-positive SUS with 20 participants we achieved a mean SUS of 78.0 (standard deviation 9.3, margin of error 4.34).

We can therefore report[10] with 97.5% confidence that the actual score for this system is above 73.66, and with 95% confidence that the actual score lies between 73.66 and 82.34.

A raw score of 78.0 has an adjectival categorization of "good" [Bangor et al. 2009] and is placed in the 82.7% percentile rank when compared to all products surveyed in Sauro [2011].

5.3 Naming exercise

Participants presented with a naming task (Figure 4) produced responses with consensus between 36.4 and 90.9% (Table 1).

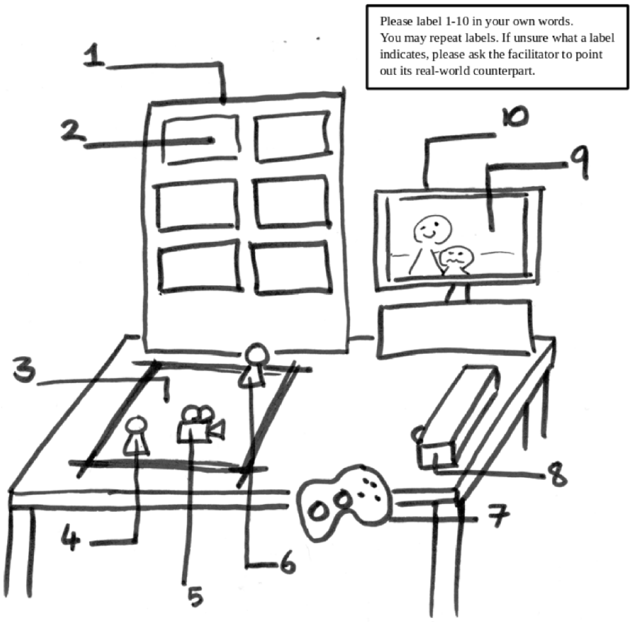
5.4 Diagnostic feedback

Diagnostic feedback from Likert scale questions is graphed (below) in Tukey boxplots[11] [Frigge et al. 1989]. These qualitative

[10] SUS data capture and calculations were made with the SUS Calculator Package v1.42 [Sauro 2011].

[11] Boxplot values (quartile boundaries and outliers) were calculated and graphed with Gnumeric Spreadsheet v1.12.8.

results have been augmented with observations of relevant participant behavior, recorded comments, “think-aloud” statements and interview statements.



Please label 1-10 in your own words. You may repeat labels. If unsure what a label indicates, please ask the facilitator to point out its real-world counterpart.

Figure 4. Image presented to participants in semi-structured naming exercise. Instruction reads "Please label 1-10 in your own words. You may repeat labels. If unsure what a label indicates, please ask a facilitator to point out its real-world counterpart."

Table 1. Comparison of original and common terms for system components and percentage consensus for common terms.[12]

Item	Original researcher term	Most common participant term	% consensus
1	Storyboard	Storyboard	81.82
2	Keyframe	Frame	36.4
		Panel	36.4
3	Tabletop	Stage	54.55
4	Model	Character	72.7
5	Camera model	Camera	90.9
6	Model	Character	72.7
7	Gamepad	Controller	90.9
8	Depth sensor	Kinect	45.5
9	Virtual scene	Scene	45.5
10	Monitor	Monitor	70.0

[Gnumeric 2013, Keeling and Pavur 2011]

5.5 Tangible measures

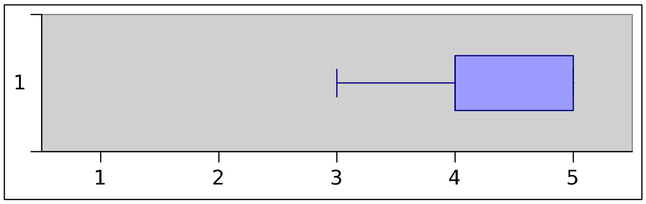


Figure 5. The size of the models made them easy to handle.

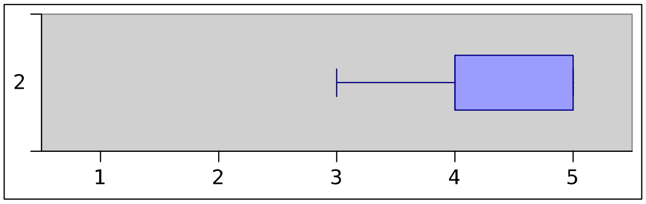


Figure 6. The stage area was big enough to accomplish the tasks easily.

Figure 7. The models were easily distinguishable from one another.

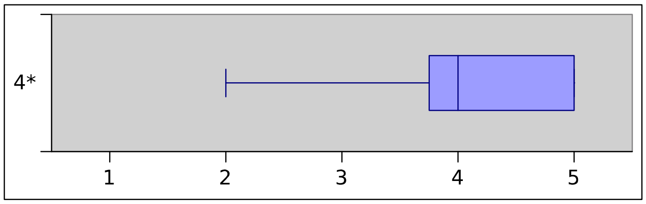


Figure 8. The models were unpleasant to the touch. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

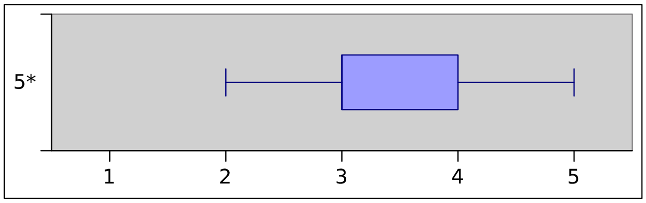


Figure 9. The models were unstable or fell over. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

---

[12] Alternative labels given for 1: pitchboard, sequence; 2 shot, storyboard; 3: board, work area; 4 and 6: macquette, object, player, proxy; 5: object, proxy; 7: gamepad; 8: black box of doom, capturer, sensor, tracker; 9: camera view, display, previz capture; 10: screen.

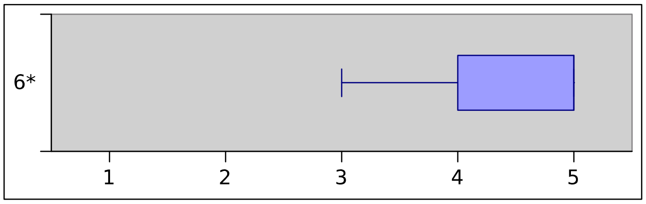


Figure 10. I was concerned about breaking the models. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

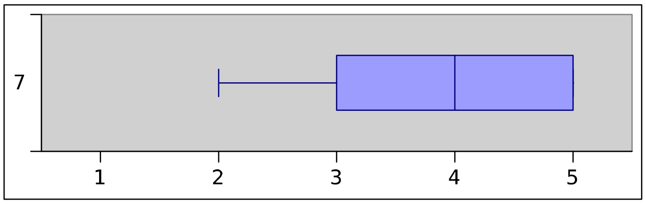


Figure 11. It was easy to imagine the other models from the model camera's point-of-view.

The results for spatial attributes (figures 5 & 6), tactile attributes (figures 5, 8, 9, 10), visual attributes (figures 7 & 11), and cognitive attributes (figure 11) all support hypothesis 2: in all cases quartile 2 ≥ 3 (Q2≥3), meaning 75% of responses were neutral-or-better.

This is corroborated by spontaneous user comments such as “This is very easy!”, “I love the tactile thing, moving things is so much faster”, “I can see things in actual 3D, get a better feel for the space”, “...such a nice representation of the characters' relative sizes.”, “I liked the hands-on creative feel of it”.

The extreme outlier (“agree” rather than “strongly agree”) for figure 7 (models distinguishable) had no attached comment. The sole negative result for figure 8 (unpleasant to the touch) had the comment “too light”. The negative responses for figure 9 (models unstable) confirmed prior observation that one model in particular fell over frequently. One of the two negative results for figure 11 (cognition) had a comment: “I felt that the camera was off-center”.

Participants engaged with the camera's point-of-view both physically and cognitively. Physically, many frequently knelt, squatted or otherwise arranged themselves to examine the scene from the model camera's line-of-sight (“I'm looking top-down to line it up with the camera.”). Cognitively, *all* participants confirmed any adjustments they made by standing upright and repeatedly turning their gaze from tabletop to screen and back again (in one case, a participant spent most of their time in this 'scanning' phase, needing to make very few actual adjustments).

Furthermore, the sense of a tangible ideal (“input=output” [Ishii 2008a, Ishii et al. 2012]) was reinforced by comments about nice-to-have features or unexpected behaviour: “it would be nice if they [the models] were pose-able” (radical atoms), “it seems to be rotating the other one even though I haven't changed it” (accuracy limitations violating the input=output ideal), “if I look at the camera's front shape here, yeah the frustrum, I keep thinking I'll see only this [indicates cone of space on stage], but it actually sees all this [makes wider cone]” (expectation that the virtual camera's field-of-view should match the implied frustrum of the physical camera model).

5.6 Gamepad measures

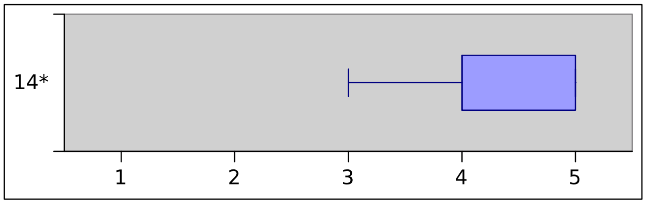


Figure 12. I had to look up the controls. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

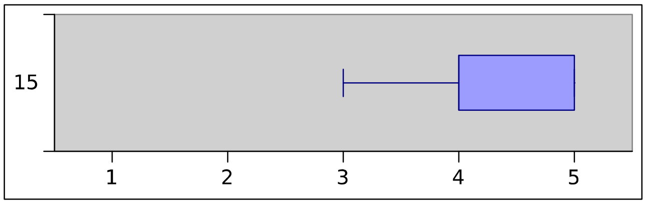


Figure 13. I felt comfortable and confident using the game controller.

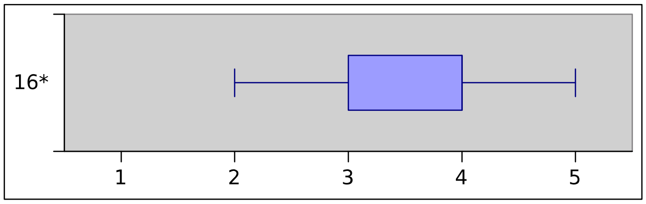


Figure 14. For this system, I would have preferred to use a mouse and keyboard. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

The results for gamepad mapping (figures 12 & 13) and appropriateness (figures 13 & 14) support hypothesis 2 (in all cases Q2≥3).

This is corroborated by positive comments (“I like using the controller”, “Only had to be explained once”) and mostly neutral-or-better scores from participants who at first expressed dislike of or low confidence in gamepads at the beginning of the session.

There was no comment on the only strong indication of mouse/keyboard preference (figure 14). This preference report came from a participant who reported low familiarity with gamepads (see Figure 15).

When plotting gamepad responses against reported familiarity (see Figure 15), it can be seen that the most negative responses for confidence and preference for gamepad over keyboard/mouse came from participants who reported low familiarity. However, the same demographic also reported high levels of confidence and preference for gamepad over keyboard/mouse.

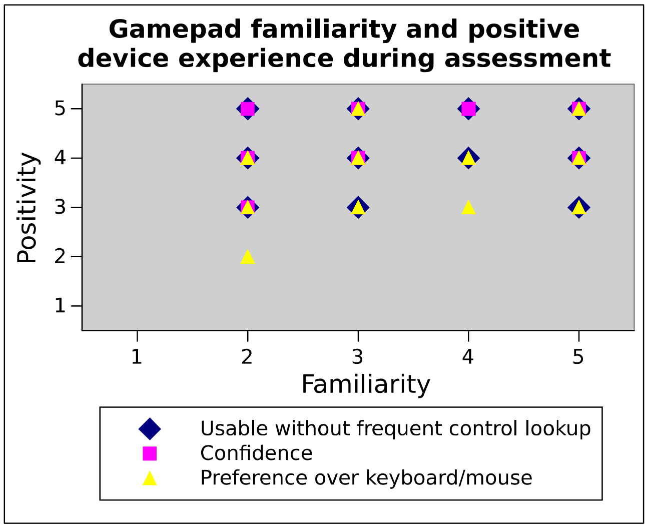


Figure 15. Gamepad usability measures against reported familiarity with gamepads. Lowest responses for confidence and preference for keyboard/mouse are associated only with low familiarity.

5.7 Pose and comfort

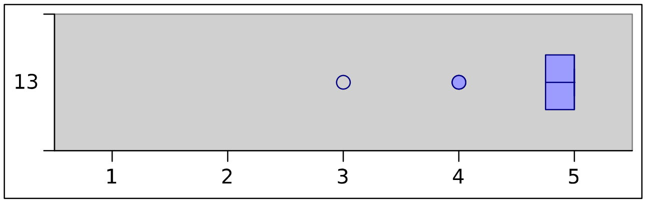


Figure 16. I found the system easy to use while sitting (from 12 responses).

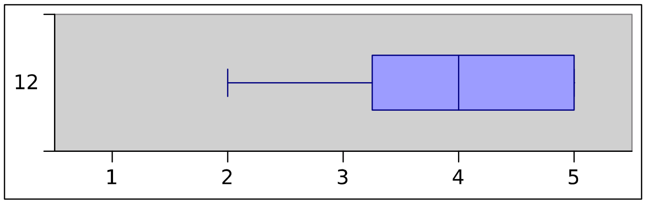


Figure 17. I felt comfortable standing for the full session (from 14 responses).

Participants were observed spontaneously adopting a number of poses while using the system, including standing, sitting and crouching. When not in use, participants tended to place the controller on the table (sometimes in view of the depth sensor) or held the controller in one hand and rested it against their body. Participants also chatted comfortably while using the system, and sometimes fielded minor interruptions (e.g. phonecalls and impromptu meetings).

14 participants reported standing while using the system, or gave a “neutral” response indicating they felt they could have. 11 participants reported sitting while using the system and 1 gave a “neutral” response indicating they felt they could have. Most participants either stood for the whole session or sat for the whole session.

Their responses support hypothesis 2 (Q2>3 for standing, Q1>4 for sitting).

5.8 Accuracy

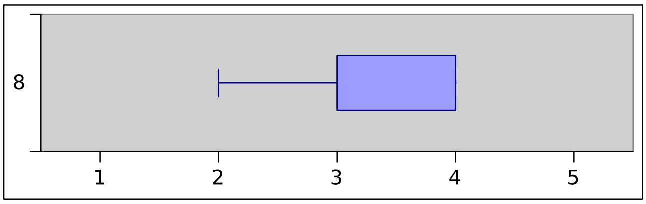


Figure 18. The accuracy of the virtual scenes created from my tabletop arrangements was [rating].

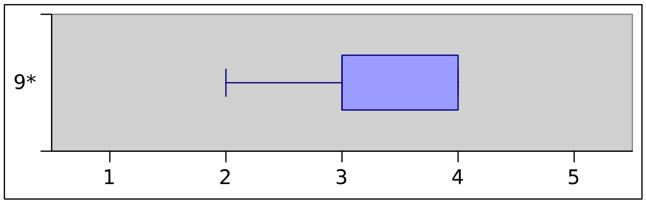


Figure 19. The system misidentified models. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

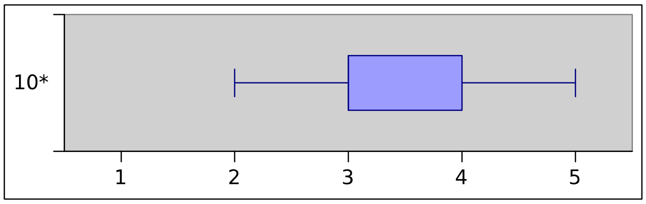


Figure 20. The system got the positioning of models wrong. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

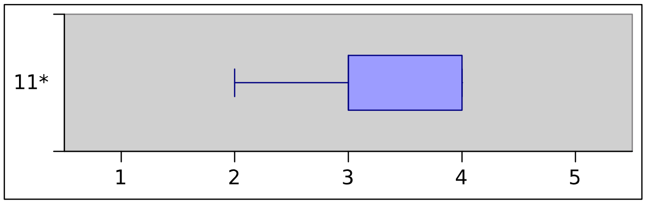


Figure 21. The system got the rotation of models wrong. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

The results for general accuracy (figure 18), model identification (figure 19) and model registration (figure 20 & 21) support hypothesis 2 (in all cases Q2≥3).

Participants reported difficulty assessing whether an error in the appearance of the virtual scene was due to misidentification or an inaccurate rotation of a correctly identified model. For example, if the scene appeared empty, it could be that the camera was rotated

180° (i.e. away from the character models), or that the camera had been misidentified by the Kinect, causing further confusion. This may have influenced responses to accuracy questions (figures 19-21). The general measure of accuracy (figure 18) may be the best representation of their sense of the system's accuracy. Despite occasional frustrating errors, 75% of participants gave a neutral-or-better assessment on all measures of accuracy.

5.9 Time

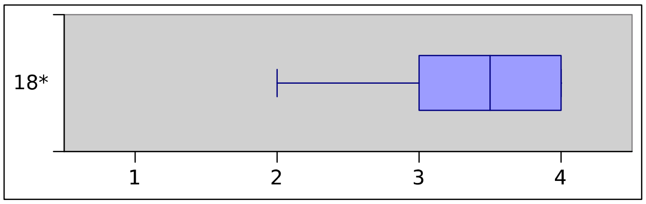


Figure 22. Usage barrier: The time taken to arrange models on the table is [rating]. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

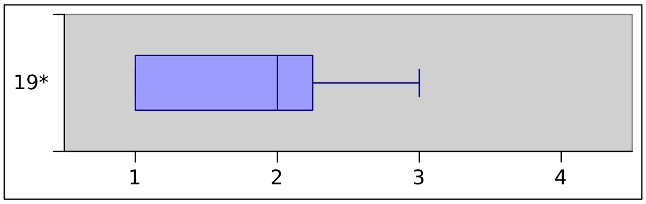


Figure 23. Usage barrier: The time the computer takes to identify the models is [rating]. (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

Time taken to perform tangible interactions did not present a barrier (figure 22), supporting hypothesis 2 (Q2≥3; note: 3 remains the "neutral" state despite this figure having a 4-point Likert scale).

However, computation time for identification and registration present a barrier to users (figure 23). This fails to support hypothesis 2 (all responses are neutral-or-worse). This result was confirmed by polite expressions of frustration ("ja, it's… taking… er, quite a while") while waiting for the progress dialog to complete, and more significantly by expressions of dismay when the accuracy was off. In post-session interviews, all participants included performance speedup in their top 3 must-haves.

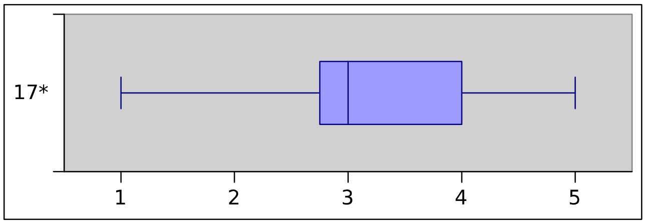


Figure 24. I would have preferred to use animation software for this task. (e.g. Blender, Maya) (Graph inverted for readability so that high values can consistently be read as supporting hypothesis 2.)

5.10 Preference for traditional software

Preference for traditional animation systems had the widest range (1-5) of all diagnostic questions. The result failed to support hypothesis 2 (Q3≥3, 50% of responses were neutral-or-worse, 50% of responses were neutral-or-better). The boxplot reveals a skew leaning towards hypothesis 2 (Q2>2.5, 50% of neutral-or-worse responses fell closer to neutral than "disagree").

When plotting preference for traditional animation systems against reported familiarity with those systems (see Figure 23), it can be seen that the strongest preference came from the strongest familiarity, and the strongest preference against came from the weakest familiarity.

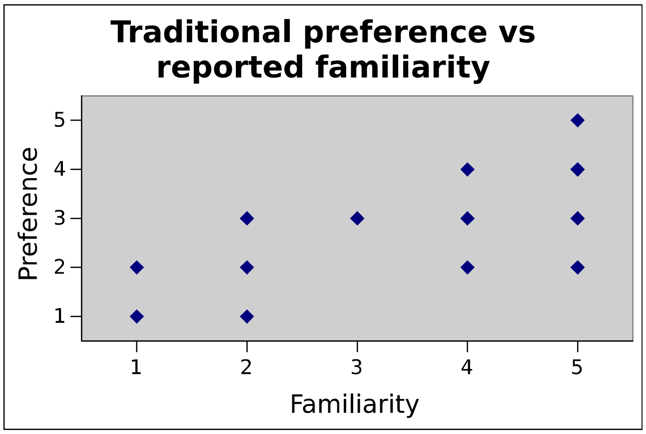


Figure 25. Preference for traditional software (5 indicates strong preference for traditional software) against familiarity with traditional software (5 indicates extremely familiar with traditional software).

6. DISCUSSION

6.1 Benchmark

The benchmark SUS score of 78.0, and associated "good" interpretation [Bangor et al. 2009], was achieved in spite of limited accuracy and poor wall-clock speed. As a lower bound on the potential usability of the system [Sauro 2011], this result suggests that future prototypes and any final implementation would likely achieve the proposed ideal [Sauro 2011] of 80.0 ("excellent"). Accuracy and speed are obvious candidates for attention.

Thus there is robust [Bangor et al. 2008, Sauro 2011] qualitative confirmation that a tangible modality can be considered usable when applied to storyboarding and scene layout tasks.

SUS measures of future iterations can be compared to this benchmark, for example to confirm that an enhancement or new feature improves usability. Meaningful comparisons can also be made with other systems targeting the same task (including expert users using traditional animation software) and with other tangible systems.

While it could be suggested that this promising score arises primarily from perceived novelty ("This was fun", "It's a break from staring at Maya on a screen all day"), the reality-based interaction framework (RBI) [Jacob et al. 2008], studies of systems with multiple modalities [Oviatt 1999] and the body of work on tangible systems in particular [Ishii 2008b] would attribute the usability to an engagement of users' spatial and body awareness. This claim could be verified by performing SUS benchmarks of other reality-based, multimodal and/or tangible systems, and comparing these to existing benchmarks from traditional interaction paradigms.

The remainder of this discussion is limited to interpretations of the qualitative results, and is focused on practical issues and the requirements of animation professionals.

6.2 Naming exercise

The naming exercise confirms that our language differed from the language used by participants, particularly for the tangible components (stage and printed characters) and atypical hardware (Kinect and controller). We had incorrectly reasoned that our lack of familiarity with animation-specific terms (e.g. storyboard, frame) had led us to mislabel those elements, causing the communication difficulties previously described. However, these results suggest that in fact our own domain- and system-specific terminology was to blame (e.g. “depth sensor”). Our conclusion is that, usability notwithstanding, any system making use of atypical hardware or interfaces should make a labeled diagram available to the user to aid communication.

6.3 Tangible measures

Responses and behavioral observations confirmed that participants engaged with the cognitive aspects of the task (consideration of the camera's point-of-view) via the tangible modality. While the task relied on developed cognitive skill for some participants (“I do this kind of brainwork regularly”), this skill was not universal and would not apply to graduate hires (“we could give this to juniors”). Even among those who expressed strong familiarity with the cognitive space, participants suggested that the system had industry-specific utility (“you could use this to take measures [a step in the scene layout task]”, “we could use this to test out shots”, “I [a director] could use this and send it to the [story]boarder and previz at the same time”).

Although aggregated tangible measures confirm of feedback from the heuristic phase of this study (model size, stage size, and other measures) these do not imply limits on future implementations. For example, one participant agreed that models were an appropriate handling size, but expressed openness to them being “half as small or 1.5 times as large”.

No models broke during the study, although breakages had occurred beforehand on thin structures such as tails (which were removed) and ankles. Participants perceived the models to be robust, which implies that in the long term their handling might be too rough for models with thin structures. In future, models should be reinforced around thin structures, without severely compromising representation.

6.4 Gamepad measures

Positive feedback from novice gamepad users (some with observed negative expectations) and expert gamepad users (some with strong habits drawn from conventions) suggests that an untethered, hand-friendly input device is a suitable “remote control” complement to a tangible system, should it need one.

More broadly, these results suggest that the selected gamepad mappings were learnable to users with at least passing familiarity with gamepads. One possible explanation is that these users had experience but lacked confidence (“but I'm not really a gamer”), had learned more than they realized by watching others use gamepads, or were assumed that a gamepad implied memorising a mapping that used all available buttons (“I always worry I won't remember which one to press”). Although no conclusions can be drawn about users with *no* prior exposure, this positive result implies that gamepads should not be dismissed as potential input devices in other non-gaming contexts.

A gamepad was selected only after technical issues with a particular model of semaphoric armband (discussed above). We note that the Myo API continues to be updated, and investigation into a semaphoric-tangible pairing remains open.

6.5 Pose and comfort

Participants were not specifically instructed to sit or stand, so the positive response to both options can be attributed to their freedom to choose the pose they preferred, and the system's capacity to support both options. While a large, accessible surface lends itself to standing operation, space is limited in a working environment, and so sitting operation an important measure of industry viability.

It is likely that having a gamepad contributed to this freedom and comfort, as suggested by the variety of ways that participants rested the gamepad on their own bodies: beyond the design intention that the device simply be easy to put down and pick up.

6.6 Time and accuracy

There was surprising tolerance reported for inaccurate results in survey results and post-session interviews. One explanation is that our participants expect a workflow where their first attempt, whether digital or in a sketchbook, must be approved by others, and is first iteration on the way to a final product. Further, they expressed satisfaction with their final frames, and attributed the need to make additional attempts to themselves rather than the system (“I'm a perfectionist”, “I just can't help myself”).

Participants were frequently observed making multiple minor adjustments to the placement of models on the stage in an attempt to achieve minute changes in the virtual scene (“Look here [points at monitor]. This guy's eyebrow is about 1 millimeter too close to the top of the frame”). When questioned about this, they reported that their work demands exacting attention to detail (“I've been asked to move a blade of grass in a shot before. Seriously. That's normal here.”).

Thus, improvements to registration accuracy are unlikely to inspire a preference for TUI over traditional animation software. Animation professionals require not only accuracy, but also extremely fine-grained control. It remains for future research to investigate whether the human hand and eye can arrange, position and rotate objects just fractions of a degree or millimeter at a time, especially under significant time pressure.

Results highlighted the impact of time on any practical implementation. While a progress dialog could not make participants ignore the time they had to wait, we believe that implementing one reduced any sense of punishment participants may have experienced. (Consider the increased frustration if participants did not know that they had to wait.) Participants still felt punishment if at the end of the wait the result was not what they wanted or intended, but gave clear feedback that the wait was the more punishing.

Finally, we note that accuracy degraded in the mid-to-late afternoon each day of the user experiment, despite artificial overhead lighting. We assume this was caused by light levels lowering by a degree insignificant to human vision, but compromising for the vision component of the registration module. We consider this a significant practical impediment to workplace viability.

6.7 Suitability to storyboarding

We can see that the tangible modality is generally suitable for layout tasks, which fits our theoretical analysis. However, this system would not be suitable for industry storyboarding. Even if time and accuracy were improved, it is not guaranteed that users' hand-eye coordination could meet their accuracy requirements (see 6.6). In addition, physical objects have an inherent inflexibility that may be too great for the diversity of scenes that might be included in an animated film (extreme close-ups, panoramic shots or shots focusing on articulated animation – all of which can be achieved by traditional sketch-based storyboarding and animatics). Finally, this inflexibility also extends to critical system actions: the tangible component cannot revert to a previous state (“No I don't like this, now I have to set it up all over again”, “In Maya I could just Ctrl-Z”), which imposes limits on users' ability to perform post-hoc adjustments without virtual assistance.

7. CONCLUSIONS

Hypothesis 1 is supported as the system achieved an SUS benchmark of 78.0. This corresponds to an adjectival rating of “good”, placing it at above-average usability. Furthermore, the benchmark provides a point of comparison for future improvements of this system, or for comparisons with similar systems.

Hypothesis 2 is not supported overall. Though the tangible components of the system are pleasant and fun to use, participants reported critical barriers to adoption. Lack of fine-grained control and flexibility were highlighted as points that led to a preference for traditional animation software for such a specific task. Similarly, core tasks like “undo” are not implementable in the tangible interface with current technology.

In conclusion, any tangible animation system would need to be augmented with a complementary modality that addresses these limitations.

8. FUTURE WORK

Further study into the limitations of human hand-eye coordination and precision is a promising next step in assessing the feasibility of tangible interfaces in the workplace. The inclusion of related issues like concentration and fatigue is essential in considering mainstream adoption given the long usage periods that come with the length of a typical work day.

There is also great scope for innovation of supplementary mechanisms for very fine rotation or translation of objects, and/or performing core system tasks. It remains to be seen the extent to which a material solution to these problems can be engineered, and whether a solely tangible solution is necessarily preferable to a multi-modal alternative.

9. ACKNOWLEDGMENTS

I would like to acknowledge and thank the following people:

- Mike Buckland and Stuart Forrest (Triggerfish Animation Studios) for hosting my study during the late 2015 campus closure
- Liza Kok (Triggerfish) for her astonishing insights and for helping me recruit sufficient participants to achieve a benchmark
- Kane Croudace (Triggerfish) for his much-needed lessons on the previz process
- The Blender Foundation and catart3d for providing the freely licensed assets used in this study
- Dr Melissa Densmore (UCT) for feedback during the prototyping phase
- Dr Hendranus Vermeulen (UCT) for educating me on the qualitative approach
- Dr Patrick Marais (UCT) for his excellent feedback
- My supervisor, Dr James Gain (UCT), for giving me considerable intellectual freedom to explore this topic, yet somehow being available for guidance whenever I found myself stuck
- My life and project partner, David Seaward, for providing the registration module, and for working so many long nights beside me.

10. REFERENCES

[1] Aaron Bangor, Philip Kortum, and James Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574-594.

[2] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114-123.

[3] Bill Buxton. 2010. Visual Storytelling. In *Sketching user experiences: getting the design right and the right design.* Morgan Kaufmann (2010), 277-297.

[4] John Brooke. 1996. SUS: A “quick and dirty” usability scale. *Usability evaluation in industry* 189, 194 (1996), 4-7.

[5] John Brooke. 2013. SUS: a retrospective. *Journal of Usability Studies* 8, 2 (2013), 29-40.

[6] Pablo Figueroa, Julian Arcos, Diego Hernando Rodriguez, Juan Manuel Moreno, and Faramarz F. Samavati. 2014. A Pen and Paper Interface for Animation Creation. *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications (GRAPP/IVAPP)*, 5-8.

[7] Michael Frigge, David C. Hoaglin, and Boris Iglewicz. Some implementations of the boxplot. *The American Statistician*, 43, 1 (1989), 50-54.

[8] Sacha Goedegebure, A. Goralczyk, E. Valenza, N. Vegdahl, W. Reynish, B. V. Lommel, C. Barton, J. Morgenstern, and T.

Roosendaal. *Big Buck Bunny* (2008) Retrieved November 1 2015 from https://peach.blender.org

[9] Gnumeric. 2013. Gnumeric 1.12.8 aka "TBD" is now available. Retrieved November 1 2015 from http://www.gnumeric.org/announcements/1.12/gnumeric-1.12.8.html

[10] Hiroshi Ishii. 2008a. Tangible bits: beyond pixels. *Proceedings of the 2nd international conference on Tangible and embedded interaction* (February 2008), xv-xxv.

[11] Hiroshi Ishii. 2008b. The tangible user interface and its evolution. *Commun. ACM* 51, 6 (June 2008), 32-36.

[12] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 201-210.

[13] Kellie B. Keeling and Robert J. Pavur. 2011. Statistical accuracy of spreadsheet software. *The American Statistician* 65, 4 (2011), 265-273.

[14] Kwegyir (Bilo) Lwabona. 2015. *AniTouch: Multitouch-based Collaborative Pre-visualisation for Computer Animation.* Master's thesis, work in progress. University of Cape Town, South Africa.

[15] Maletsabisa Molapo and Gary Marsden. 2013. Content Prototyping – An Approach for Engaging Non-technical Users in Participatory Design. *Human-Computer Interaction (INTERACT 2013)*, 788-795.

[16] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* (1990), 249-256.

[17] Jakob Nielsen. 1995. 10 usability heuristics for user interface design. 1 January 1995. Retrieved November 1 2015 from http://www.nngroup.com/articles/ten-usability-heuristics/

[18] Donald A. Norman. 2010. Natural user interfaces are not natural. *interactions* 17, 3 (May 2010), 6-10.

[19] Kenton O'Hara, Richard Harper, Helena Mentis, Abigail Sellen, and Alex Taylor. On the naturalness of touchless: putting the “interaction” back into NUI. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 1 (2013), 5.

[20] Sharon Oviatt. 1999. Ten myths of multimodal interaction. *Commun. ACM* 42, 11 (1999), 74-81.

[21] Joshua G. Tanenbaum, Amanda M. Williams, Audrey Desjardins, and Karen Tanenbaum. 2013. Democratizing technology: pleasure, utility and expressiveness in DIY and maker practice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), 2603-2612.

[22] Marc Rettig. 1994. Prototyping for tiny fingers. *Commun. ACM* 37, 4 (1994), 21-27.

[23] Ton Roosendaal. 2015a. Blender 2.73 - A New Storyboard Workflow. *Cosmos Laundromat – The Gooseberry Open Movie Project* (January 15 2015). Retrieved November 1 2015 from https://gooseberry.blender.org/blender-2-73-a-new-storyboard-workflow/

[24] Ton Roosendaal. 2015b. Blender 2.8 - the Workflow release. *Blender developer blog* (July 20 2015). Retrieved November 1 2015 from https://code.blender.org/2015/07/blender-2-8-the-workflow-release/

[25] Ton Roosendaal. 2015c. Blender 2.8 - the Workflow release. *Bf-committers mailing list* (July 26 2015). Retrieved November 1 2015 from http://lists.blender.org/pipermail/bf-committers/2015-July/045721.html

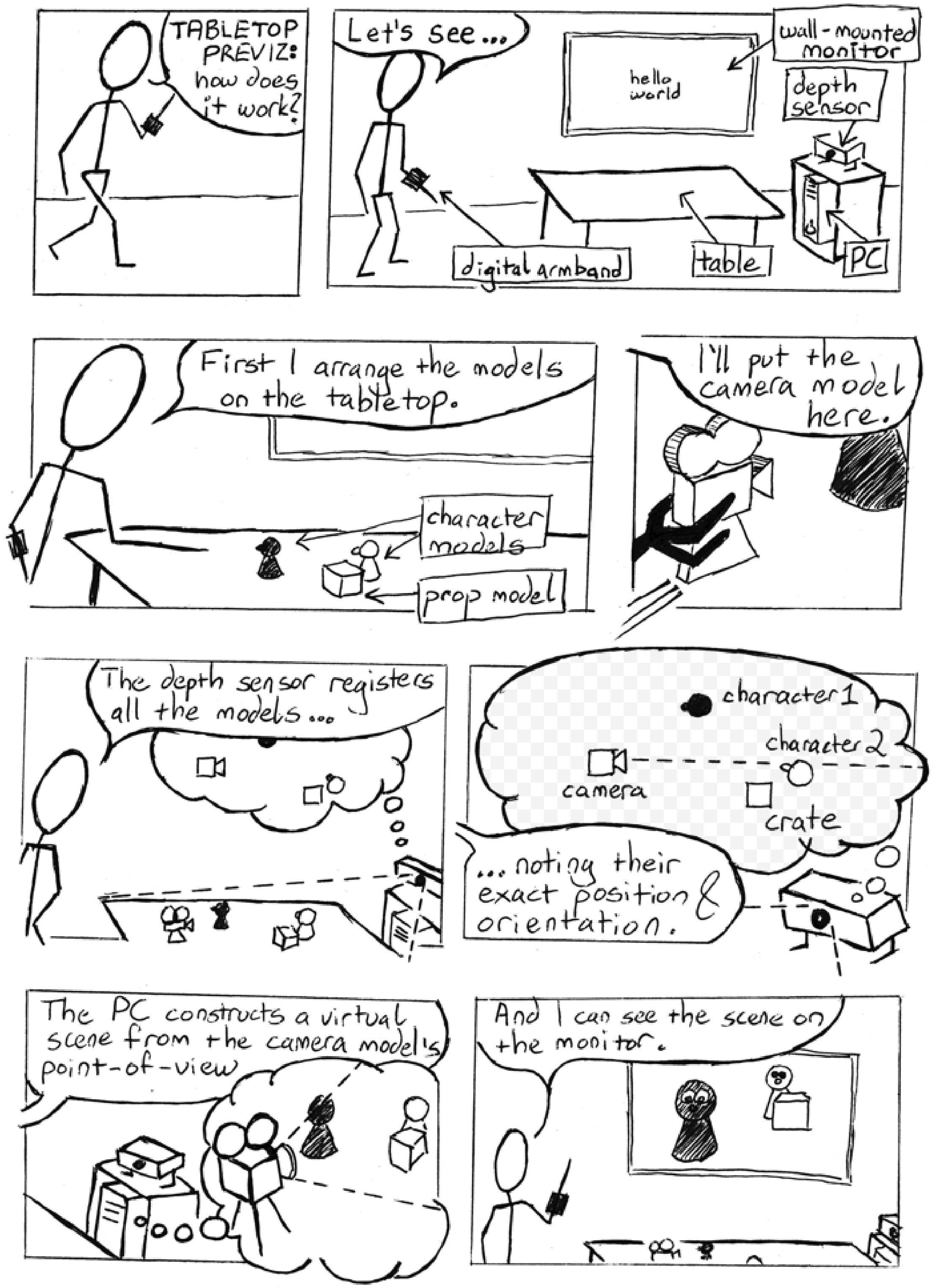
[26] Jeff Sauro. 2010. Does better usability increase customer loyalty? *MeasuringU* (January 7 2010). Retrieved November 1 2015 from http://www.measuringu.com/usability-loyalty.php

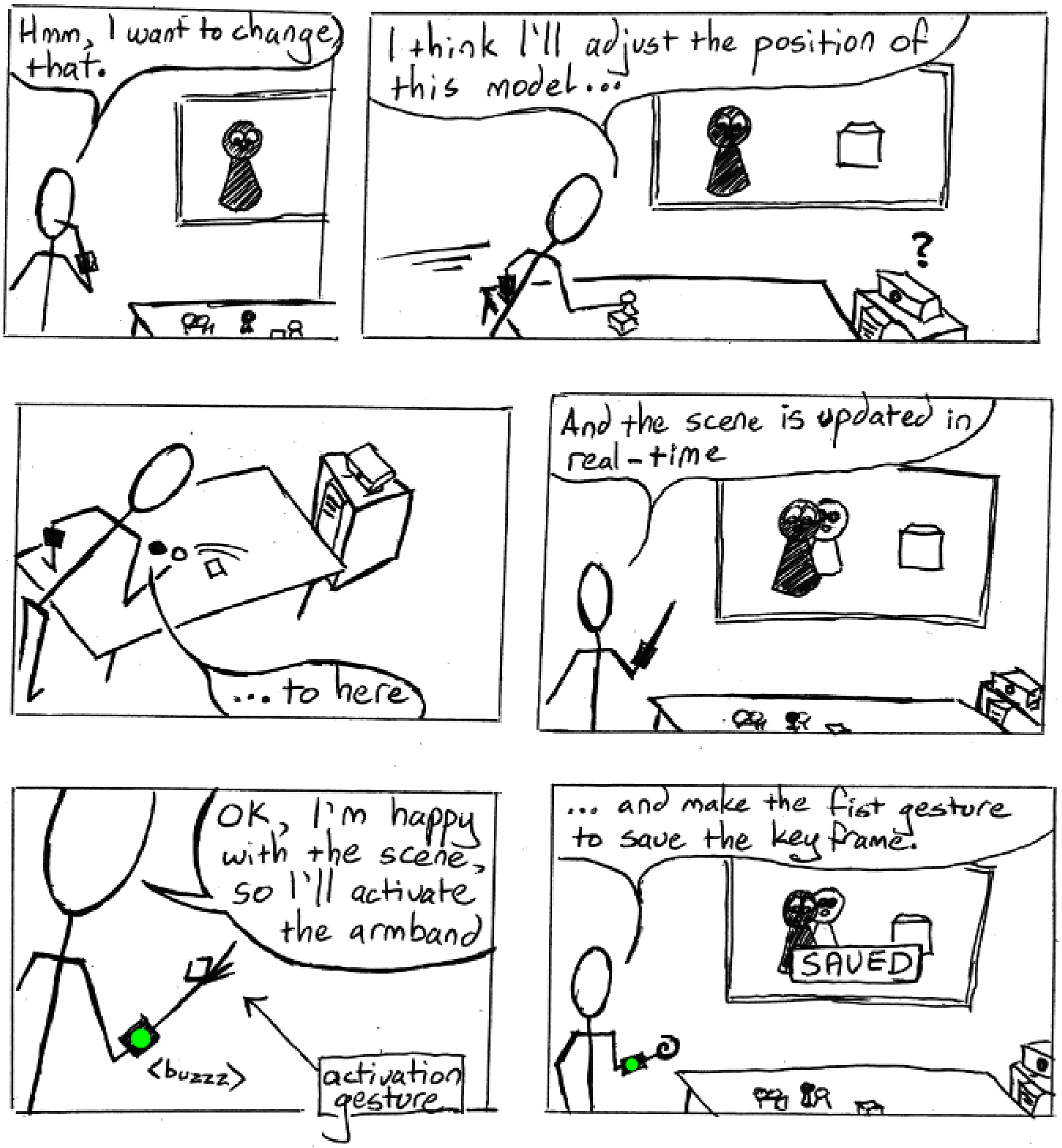
[27] Jeff Sauro. 2011. *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices.* Measuring Usability LLC (2011).

[28] Jeff Sauro and James R. Lewis. When designing usability questionnaires, does it hurt to be positive? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), 2215-2224.

[29] Orit Shaer and Eva Hornecker. 2010. Tangible user interfaces: past, present, and future directions. *Foundations and Trends in Human-Computer Interaction* 3, 1-2 (2010), 1-137.

Appendix A: Illustrative comic





Tabletop previz: how does it work?

as part of "PREVIZ 2015"