Engaging Families around Museum Exhibits: Comparing Tangible and Multi-touch Interfaces

Michael S. Horn[1,2], Amartya Banerjee[1], David Bar-El[2], Izaiah Hakim Wallace[1]
Computer Science[1], Learning Sciences[2]
Northwestern University
Evanston, IL, USA
michael-horn@northwestern.edu, amartyabanerjee2012@u.northwestern.edu,
davidbarel2021@u.northwestern.edu, izaiahwallace2014@u.northwestern.edu

ABSTRACT
Over the last decade, large multitouch displays have become commonplace in museums and other public spaces. While there is preliminary evidence that exhibits based on tangible technologies can be more attractive and engaging for visitors than displays alone, very little empirical research has directly compared tangible to large multitouch displays in museums. In this paper, we present a study comparing the use of a tangible and a multitouch tabletop interface in an exhibit designed to explore musical rhythms. From an observation pool of 791 museum visitors, a total of 227 people in 82 groups interacted with one of the two versions of our exhibit. We share the exhibit design, experimental setup, and methods and measures. Our findings highlight advantages of tangible interaction in terms of attracting and engaging children and families. However, the two exhibits were equally effective at supporting collaborative interaction within visitor groups. We conclude with a discussion of the implications for museum exhibit design vis-à-vis visitor engagement and learning.

Author Keywords
Tangible interaction; museums; informal learning; exhibit design; interactive tabletops; multi-touch displays.

CCS Concepts
• Human-centered computing~HCI design and evaluation methods • Human-centered computing~Interaction paradigms

INTRODUCTION
Since the 1960s, interactive exhibits have become a cornerstone of the museum experience [24]. Well-designed exhibits can result in rich and rewarding learning for children and adults by supporting meaningful conversation [4, 5, 6, 12], experimentation and discovery [16], and enactment of scientific practices [23]. It is a way to touch and explore subject matter in a low-pressure, open-ended space where people can make meaning of experiences with family and friends. As Frank Oppenheimer, founder of the Exploratorium in San Francisco, put it, "No one ever flunked a museum." [24].

Designing high quality interactive exhibits is challenging. It requires a consideration of how to invite visitors of different ages, genders, and experiences to engage with and around the exhibit. The design materials used in exhibits shape how parents talk to their children about scientific objects [6]. Moreover, exhibits can build on cultural forms and allow visitors to draw on their existing practices [12]. Studies have also shown that good design decisions can lead to more in-depth engagement and repeated play with an exhibit [16, 37]. As different technologies become available in museums, it is important to re-examine how materials lead to behaviors associated with museum learning.

Over the last decade, large multi-touch displays have become commonplace in museums and other public spaces. These displays can be used in conjunction with exhibited objects [31], or serve as exhibits in their own right (e.g. [4, 13, 20, 37]). Somewhat less common are exhibits that make use of tangible technologies and interactions. By tangible [32, 36], we refer to exhibits that feature physical objects and/or whole-body interactions [33] to control digital systems. Researchers have reported on some advantages for tangible interfaces [14, 21, 37], with most focusing on visitors'

Figure 1: Family interacting with the tangible version of our exhibit at the science museum.

IDC '20, June 21–24, 2020, London, United Kingdom
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7981-6/20/06…$15.00
https://doi.org/10.1145/3392063.3394443

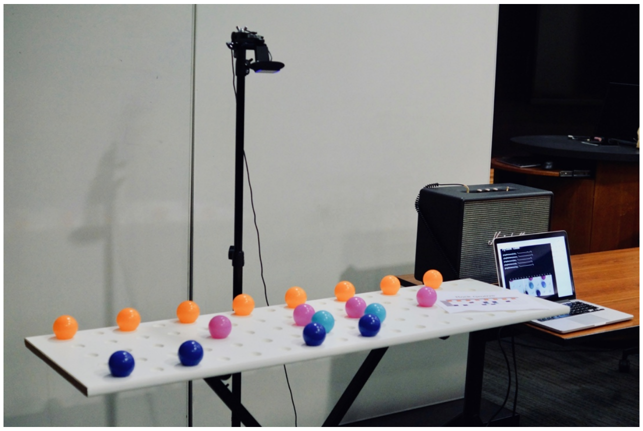


Figure 2: Tangible version of the exhibit. Laptop drives the connected speakers to give instantaneous A/V feedback based on the beat pattern.

manual actions—how visitors use their hands to manipulate the digital system with and without tangibles.

While using a visitor's manipulation of the exhibit as the unit of analysis is important, we argue that analyzing visitor behavior around the exhibit is beneficial as well, especially in terms of attracting and engaging visitors in an equitable way. By documenting visitors' behavior as they walk past and stand by the exhibit, we capture the degree to which the two technologies attract and engage visitors of different backgrounds.

In this paper, we report on a study in a science museum comparing visitor behavior around multi-touch and tangible versions of a beat-making exhibit designed to explore musical rhythm. In the first version of the exhibit, visitors could create their own beats by arranging colorful plastic balls on a white board (Figures 1, 2). In the second version, we presented a similar interface on a large multi-touch display (Figure 3, right). We observed a total of 791 people in 356 visitor groups as they came into the vicinity of the exhibit. A total of 227 people in 82 groups interacted with one of the two versions of the exhibit.

After discussing related work, we share the exhibit design, the experimental setup, and the methods and measures. We then present quantitative results, along with more detailed qualitative observations of visitor interaction. Our findings offer evidence of the drawing power of tangible exhibits, especially for groups with children, and informs museum exhibit design for visitor engagement and learning.

RELATED WORK

Museum exhibits—especially in interactive science museums—should encourage self-directed learning and allow for sense making through hands-on and mind-on exploration [5, 6, 7, 16]. In other words, letting visitors tinker is usually a primary design goal [16, 20, 24]. Our research builds upon prior work on exhibit design for informal science settings with an emphasis on tangible user interfaces (TUIs). With museum environments changing due to a proliferation of interactive displays, we discuss prior research encompassing GUI and TUI based exhibits, and briefly review their pros and cons.

Multi-touch Interactives

Multi-Touch interactives are common in public spaces [25] including museums and art galleries. They have been shown to promote engagement in both formal and informal learning environments [9, 10, 11, 13, 32]. However, prior research has highlighted certain issues, especially pertaining to group and social dynamics in multi-touch interactions. For example, multi-touch interactives in public spaces can be frustrating for groups with a large variation in visitor age; adults predominantly prefer concentrating and taking in content, while children try to continuously interact with the screen [11, 25]. Despite the prevalence of multi-touch devices in everyday life, adult museum visitors also seem to perceive multi-touch interactives to be designed primarily for children [15].

Researchers have also studied ways to harness learners' existing social practices to facilitate multi-touch tabletop interactions [2, 13], but those very practices have their own shortcomings. For example, people often place objects on an interactive tabletop much like they would with a normal "non-interactive" surface [11]; a practice that leads to issues with accidental input [15] or occlusion of graphical elements.

Comparing TUIs and GUIs

While large multi-touch and tangible interfaces both afford multi-user, bimanual interactions, a growing body of research shows advantages for tangibles for tasks that involve target acquisition, manipulation, and data exploration [1, 2, 8, 21, 34].

Antle and Wang [2] compared a tangible and a multi-touch interface for a puzzle solving task. They found that the tangible interface enabled more efficient and effective motor-cognitive strategies that could be attributed to the 3D manipulation space and eyes-free tactile feedback.

Al-Megren and Ruddle [1] conducted an experiment where university students explored data visualizations in the field of genetics with either a multi-touch or a tangible interface. Participants performed their tasks significantly faster and more efficiently in the tangible condition. Joyce et al. [21] compared two versions of a museum exhibit designed to let visitors explore the distribution of phytoplankton in different areas of the oceans—one version used tangibles rings, while the other used virtual rings to explore this large dataset. The researchers reported that the two conditions drew similar demographics in terms of gender and age group. However, the tangible exhibit yielded longer holding times. Touching the tangible rings was a precursor to prolonged engagement, and these tangible rings were better than virtual rings at prompting an initial interaction.

With museum visits often being a family occasion, researching the efficacy of TUIs in engaging groups with children is important. Xie et al. [37] studied children playing

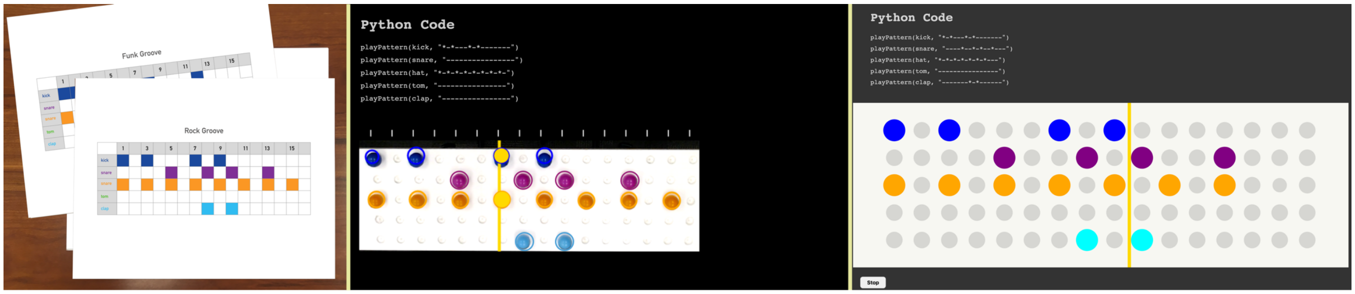


Figure 3: (Left) The beat cards meant to ease visitors into tinkering, each row corresponds to a drum sound and the 16 columns correspond to a 16th note along the step sequencer, (middle) the interface for the tangible exhibit (displayed on a laptop screen), and (right) the touchscreen/tabletop interface, very similar to its tangible counterpart.

a puzzle game with either a GUI or TUI interface. They found that the tangible interface was easier to use, and that children engaged in repeated play in the tangible condition. Such prolonged engagement is a key outcome of good museum exhibit design [16].

Beyond usability, TUIs may also be more inviting to children and promote collaboration [14, 27]. In a study conducted in a pediatric clinic waiting room, Leong [19] developed a tangible "museum" exhibit on sickle cell disease that was dramatically more engaging for young children than a tablet-based interactive.

Horn et al. developed Tern [14], a tangible programming interface, to control a robot on display at a science museum. The researchers compared a GUI implementation to the tangible version, and their results highlight that children were more likely to engage with the exhibit if it included tangibles. Moreover, the tangible exhibit facilitated active collaboration, wherein multiple members of a group were using the exhibit simultaneously.

However, the GUI in Horn et al's study was a mouse-based interface. Given that a mouse generally affords single-point, single-user interactions, the results may not be solely attributed to a lack of tangibility. In more than a decade since the publication of Horn et al., the question of whether the same effects would hold if the computer mouse condition had instead been a large interactive display, warrants further investigation. And while we have cited many scenarios where tangibles seem to have an edge—e.g. tangibles seem especially beneficial to children, who are a target demographic for museums—we also need further evidence to assess whether tangible exhibits can be more inviting than GUI exhibits.

EXHIBIT DESIGN

For this study, we created two simple temporary museum exhibits, one using a large multitouch tabletop display and one using tangible objects. We crafted the two exhibits to be as similar as possible in terms of size, color, shape, and function. Both exhibits offered visitors the opportunity to experiment with a drum machine to program musical rhythms. Programmable drum machines such as the Roland TR-808 (from the early 1980s) have significantly influenced the development of digital music and hip hop in particular. At its core, a drum machine allows musicians to program a rhythm sequence by queuing percussion sounds to play at different time steps over the course of a musical measure that loops indefinitely. Most drum machines offer a high degree of customization, but the basic elements usually consist of a four-beat measure subdivided into 16th notes or smaller intervals (Figure 4). There are also usually several tracks representing different percussion sounds.

Our exhibit design process was guided by a principle of extreme minimalism. What was the smallest feature set we could offer that would still result in an engaging visitor experience? This turned out to be a good strategy as our first pass with a minimal set of features was surprisingly engaging for visitors across a wide range of ages and group types.

Tangible Exhibit

This study is part of a larger design-based research project to engage middle school students in computational thinking through music making with computational tools. To create the tangible exhibit, we started with piece of white, laminated particle board (135cm x 38cm) and drilled an array of 16x5 holes 3cm in diameter (Figure 2). The holes were separated by 76mm center to center. Each column represented a 16th note interval in a four-beat measure with a 4/4 time signature. Each column had five holes to allow for the placement of multiple percussion sounds at each time step. We purchased a set of brightly colored plastic balls approximately 5cm in diameter. There were 80 balls total (16 in each of the 5 colors) corresponding to different percussive sounds.

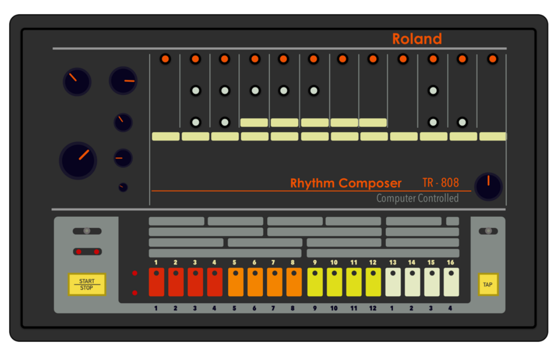


Figure 4: A four-beat measure is subdivided into 16th notes (or smaller) intervals in a Roland TR-808 drum machine.

- Dark Blue = Kick Drum
- Cyan = Clap
- Magenta = Snare
- Orange = Hi-Hats
- Green = Tom

We placed the balls in a plastic bin next to the exhibit for easy access by visitors. We used a web camera mounted approximately 1m above the exhibit to capture a video feed of the table. A laptop computer powered the exhibit. Our software was implemented in JavaScript, HTML, and Web Audio. We used the tracking.js [35] library to recognize different colored balls in the video feed in real time. We used the Web Audio library to queue one measure of audio at a time. Visitors' changes were shown immediately on the video feed (visitors could see the laptop screen) and updated in the next measure of audio. The tempo was fixed at 90 beats per minute, so one measure refreshed every six seconds. We drew colored circles around each ball on the video feed to help visitors understand how the exhibit worked, and it was common for visitors to talk about how the camera recognized the balls, and to experiment with covering balls to block them from the camera's view. We also drew a gold line that swept across the video feed in a loop show the current notes being played (Figure 3, middle).

Multi-Touch Tabletop Exhibit

The tabletop display exhibit used a 60cm x 100cm 3M projected capacitive multi-touch display. The display consisted of a white rectangle covering the width of the display with the same aspect ratio as the tangible exhibit. The white rectangle had an array of 16x5 small grey circles meant to look similar to the tangible board (Figure 3). When tapped, the grey circles would turn into larger colored circles meant to look like the tangible balls. Tapping on a ball would toggle it back to a grey circle. Each tap would also play a preview of the sound that colored circle would play when added to the beat. Unlike the tangible exhibit, each row of circles corresponded to one of the circle colors/drum sounds. We implemented this version of the exhibit using JavaScript, HTML, and Web Audio. Similar to the tangible exhibit, a gold line swept across the table from left to right to show the current notes being played (Figure 3, right).

Both versions of the exhibit also displayed Python code that was generating the beats, as an abstract, simple function call (see Figure 3) followed by a string pattern representing the active 16th notes. This code was updated in real-time as and when visitors made any changes to the exhibit. Our goal was not to teach programming, but rather to convey the idea that computer code was behind the beat generation. We were also interested in whether a simple cue like this could trigger conversations among the visitors that related to coding.

Beat Cards

Neither of the exhibits had instructions, nor did we expect most users to have a background in sound design or beat making. To ease visitors into tinkering, we placed laminated beat cards next to each exhibit (Figure 3, left). These beat cards contained rhythm patterns for programmable drum machines from various genres of music. These patterns were derived from Bardet's Drum Machine Pattern's book [3] and were simplified to only use the five percussive sounds we had selected for the exhibits.

RESEARCH QUESTIONS

Through this study we sought to answer the following research questions.

RQ1 [Holding Time] Will a tangible exhibit engage museum visitors longer than a comparable exhibit based on a large interactive display?

Based on prior work [2, 14] we assumed the answer would depend on the success of the exhibit versions in actively engaging more than one member of a visitor group and the ability to engage children. Our hypothesis was that the tangible exhibit would have a slight advantage.

RQ2 [Capture Rate] Will the tangible exhibit attract more museum visitors than the interactive display?

Based on the notion of cueing forms [12] we hypothesized that the tangible exhibit would be more intriguing and appealing to visitors, especially children. In considering capture rate, we were also interested in average group size and the age distribution of family groups.

RQ3 [Engagement Behaviors] Will the tangible exhibit foster more behaviors indicative of enjoyment or engagement?

Specifically, we coded for behaviors such as capturing photos or videos with a phone, dancing or singing, using

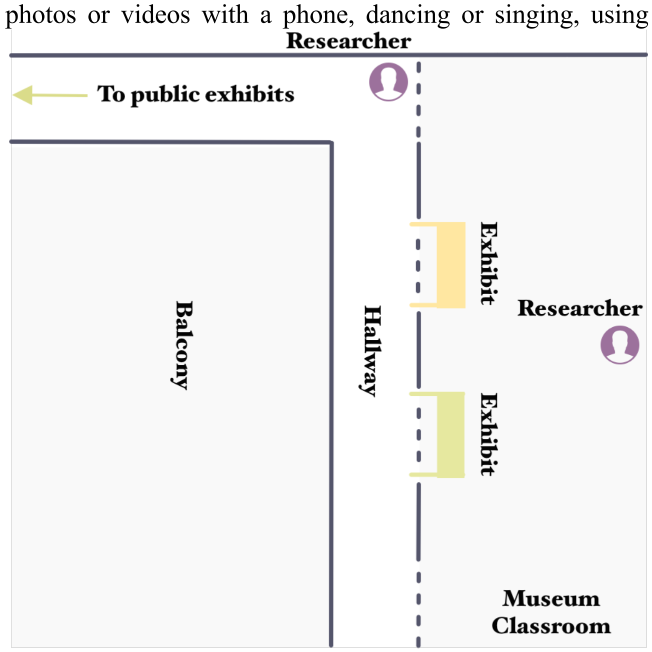
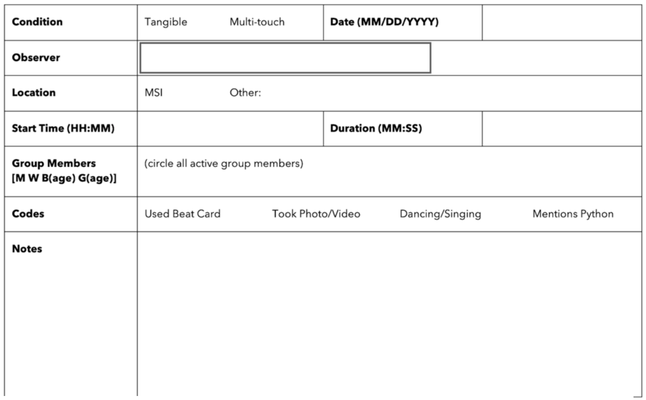


Figure 5: Exhibit layout in the museum. Sliding doors in front of the exhibits allowed us to quickly switch which exhibit – tangible or tabletop – was made available to the museum visitors.



Condition	Tangible Multi-touch	**Date (MM/DD/YYYY)**	
Observer			
Location	MSI Other:		
Start Time (HH:MM)		**Duration (MM:SS)**	
Group Members [M W B(age) G(age)]	(circle all active group members)		
Codes	Used Beat Card Took Photo/Video Dancing/Singing Mentions Python		
Notes			

Figure 6: Observation sheet (observer name is anonymized for this paper) used to collect and codify visitor data.

rhythm cards, and mentioning the python code on the screen. We hypothesized that the tangible exhibit would be more successful due to our speculation that more children would be involved and that colorful plastic balls would serve as a cueing form signaling playful engagement and fun. Note, we expected results for this question would be related to overall holding time and the number of active group members.

METHODS AND MEASURES

We conducted the study at a science museum with an audience of over 1 million visitors a year located in a large city in the United States. The exhibits were displayed in a classroom / workshop area accessible by a hallway adjacent to several public exhibits (see floor plan in Figure 5). Despite its location in a back corner of the museum, the hallway sees a substantial amount of visitor traffic. We placed the exhibits in double sliding doorways that opened out from the classroom into the hallway. We only displayed one version of the exhibit at time. For the version of the exhibit not in use, we closed the doorways to conceal it from visitors.

One researcher sat at the end of the hallway to keep track of all visitors passing through the hallway, whether or not they stopped at the exhibit, and whether or not the exhibit was currently occupied. Another researcher sat in the classroom and unobtrusively observed the exhibit keeping track of holding time, visitor behavior, and group composition. A third researcher also observed the exhibit on the first day of observations to establish inter-rater reliability. Every 45 minutes we swapped the exhibits by closing one set of doors and opening another. We also switched which exhibit was displayed in which doorway to minimize the differences the location along the hallway might have had on whether or not visitors decided to stop.

We observed the exhibit on three busy weekdays in late August 2019. The first day we fixed several technical glitches, collected pilot data, and refined our observation protocols. The following week we collected the data presented in this paper on a Monday and a Friday.

Capture Rate

To measure capture rate, the researcher sitting in the hallway recorded all visitors who passed by the exhibit. Visitors were recorded as W (adult woman), M (adult man), G (girl under 18 years old), B (boy under 18 years old). Visitor groups were recorded together on the same line of the observation sheet. For each group we noted whether or not they stopped at the exhibit, meaning that one member of the group physically touched the exhibit, even for a moment. We also noted whether or not the exhibit was already in use when the visitor group passed by. Note that all age and gender observations were estimates by the researcher (see below for inter-rater reliability). We also made judgment calls about which individuals belonged together as part of a visitor group. This was based on whether or not they were talking with one another and how close together they walked (see Block et al. [4] for a discussion of the challenges of identifying distinct visitor groups). Because we collected data in late summer there were few to no school field trips or summer camp programs, which tend to be more active earlier in the summer).

Hold Time and Engagement Codes

For the other measures, researchers in the classroom used an observation sheet (Figure 6) to record visitor group composition, holding times, number of active visitors in each group (those who physically touched the exhibit in any way), and engagement codes.

RELIABILITY

To establish inter-rater reliability, two researchers observed the exhibit in both conditions for approximately 3 hours on the first day of observations. This time period included 23 visitor groups who used the exhibit (13 groups in the tangible condition, and 10 groups in the tabletop display condition). This represents 28.0% of the total number of visitor sessions observed for the entire study.

The researchers agreed that a visitor group was present 95.7% of the time. There was only one visitor group that was recorded by one of the researchers and not the other. This group consisted of a single individual who touched the exhibit only momentarily in the tabletop display condition before moving on.

The two researchers recorded holding times within 30 seconds of one another in all but one case. Their holding times were within 15 seconds of one another in all but 3 cases. On average, the holding times recorded by the two researchers deviated by 9.91 seconds (SD=7.97).

The “Dancing or Singing” code and the “References Python” code were used very infrequently. The kappa values for these

Code	**Agreement (%)**	**Kappa**
Used Beat Card	82.61%	0.654
Photo or Video:	91.3%	0.747
Dancing / Singing	86.96%	0.330
Python:	82.61%	0.489
Any Event:	91.30%	0.823

Table 1. Inter-rater reliability values for engagement codes.

	Tangible	Multi-touch	Total
Women	145	148	293
Men	134	129	263
Girls	60	47	107
Boys	66	62	128
Total	405	386	791
Group Count	181	175	356

Table 2: Number of visitors who walked through the hallway while we were observing (by age, sex, and condition). Note, not all of these visitors stopped to try the exhibit.

two codes were too low to be considered reliable (Table 1). We report the frequency of these codes in the Results section, but we flag these results as being unreliable.

There were three groups in which the count of adult men was off by one person. On average, the researchers deviated by 0.13 individual per group.

The researchers agreed on the number of women in each group in all but two cases. On average they deviated by 0.13 women per group.

The researchers agreed on the number of boys (under 18 years old) in each group in all cases.

The researchers agreed on the number of girls (under 18) in each group in all but three cases. In one case, one researcher recorded the participant as an adult woman, and the other recorded the same participant as a 16-year-old girl. On average they deviated by 0.13 girls per group.

Putting all of these together, there was disagreement on the total number of participants in 6 out of 23 visitor groups. On average the researchers deviated by 0.30 people per group. This level of disagreement was higher than we expected. Because the researchers were positioned inside the classroom, they might have had a hard time seeing group members who were not actively involved but were observing from a few feet away.

There was also some disagreement on the number of active visitors per group. In 6 of the 23 visitor groups the researchers recorded a different number of active visitors. On average the researchers deviated by 0.348 people per group.

PARTICIPANTS

We observed a total of 791 people in 356 visitor groups as they came into the vicinity of the exhibit (Table 2). Of these, a total of 227 people in 82 groups interacted with one of the two versions of the exhibit. In the tangible condition there were 124 people in 43 groups (2.62 average group size, SD=1.14); and in the graphical condition there were 102 people in 39 groups (2.88 average, SD=1.48). The difference in average group sizes was not statistically significant.

RESULTS

Capture Rate

From a total of 181 groups in the tangible condition, 31 groups stopped to try the exhibit. For the multitouch condition, 34 out of a total of 175 groups stopped. That equates to a raw capture rate of 17.13% and 19.43% respectively. However, groups tended to spend longer at the tangible exhibit, meaning that it was occupied more often than the multitouch exhibit, which might affect capture rates. To account for this, we also noted whether the exhibit was vacant when a group passed by. When the exhibit was vacant 42.86% of groups stopped for the tangible condition, and 39.02% stopped for the multitouch condition. Interestingly, the tangible condition has a higher capture rate (51.72%) compared to the multi-touch condition (42.86%) when a visitor group included children *and* the exhibit was vacant. However, none of the differences in capture rates were statistically significant. Table 3 summarizes the capture rates at various levels of detail.

Number of Children

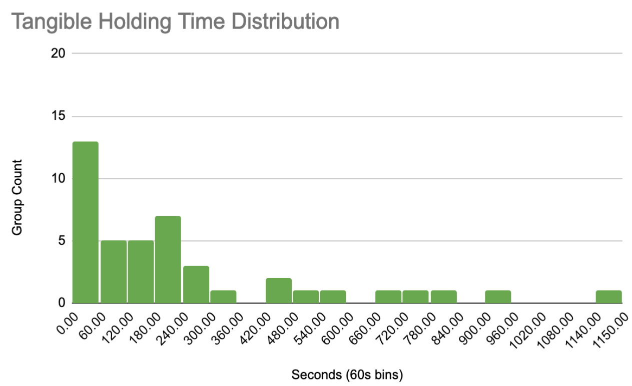
We now consider only participants who stopped at the exhibit. In this sample there were 81 children (boys or girls under 18 years old). In the tangible condition there were 52 children total with an average of 1.21 children per group (SD=1.41). 55.81% of tangible condition groups had children. In the multitouch condition there were only 29 children with an average of 0.74 children per group (SD=0.91). 48.72% of multitouch condition groups had children. A one-tailed *t*-test showed the difference in number of children to be significantly different between the two conditions (p=0.041).

Holding Time

We measured holding time as the time the first visitor in a group touched the exhibit to the time the last visitor in a group stopped interacting and walked away. Note that some of the difficulties identifying individual visitor groups

	Tangible	Multitouch
Group count	181	175
Groups with children	71	71
Number of groups that stopped	31	34
Capture rate	17.13%	19.43%
Capture rate with children	22.54%	22.54%
Stopped if vacant?	30	32
Skipped if vacant?	40	50
Capture rate if vacant	42.86%	39.02%
Stopped with children & vacant	15	15
Skipped with children & vacant	14	20
Capture rate with children & vacant	51.72%	42.86%

Table 3: Capture rates for the two conditions based on whether the exhibit was occupied and whether a group included children.



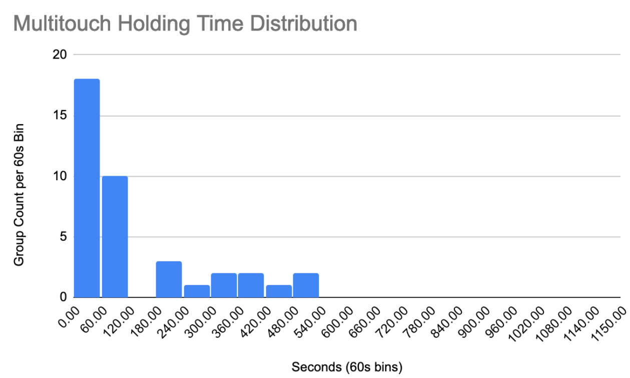


Figure 7: Holding times in 60s bins for tangible and multi-touch conditions.

described in Block et al. [4] were present here; however, because visitors were only able to access the exhibit from one side within a door frame, visitor groups for the most part seemed to take turns rather than overlapping. We recorded the start times as hh:mm (hour:minute) and then set a stopwatch to determine the number of seconds groups stayed. As noted above researchers deviated by about 10s on average in their measurements. The distribution of holding times in 60s bins are shown in Figure 7 for both exhibits.

Visitor groups in the tangible condition spent an average of 239.30 seconds (SD=272.0) at the exhibit, while visitor groups in the multitouch condition stayed for 127.77 seconds (SD=151.84), almost two full minutes less on average. A two-tailed *t*-test showed this difference to be statistically significant ($t = -2.32$, $p = 0.023$).

Interestingly, groups with children stay 100s more on average *regardless of the condition*. Similarly, groups with more than one active participant stay almost 200s more on average, reinforcing findings from [8].

Number of Active Group Members

As we were observing family groups, we kept track of *active* participants, which we defined as any group member who touched the exhibit or rhythm cards, one or more times. In the tangible condition, there were 2.12 active group members on average (SD=1.35), while there were 2.08 active group members on average in the multitouch condition (SD=1.01). This difference was not statistically significant. Because there were slightly more members per group in the tangible condition, we also looked at the percentage of active members out of the total number of people in the group. Here, the multitouch condition had a greater portion of active group members (82.78% on average) compared to the tangible condition (76.62% on average). This difference was also not statistically significant.

In Horn et al. 2009 [14], there was no statistically significant difference in holding time, but there was a significant difference between groups with only one participant and groups with multiple participants. In their study, the tangible condition had significantly more active participants. Our results corroborate some of these results. It seems that multitouch tabletops are equally good at supporting multi-user interaction and active participants.

Behavior Codes

The numbers for the behavior codes are shown in Table 4. Of those, the tangible version saw a substantially higher percentage of groups taking photos (20.93% to 7.69%) or using the beat cards (39.53% to 25.64%). As already mentioned in the inter-rater reliability section, two of the codes (Python, and Dancing/Singing) have low reliability measures based on Cohen's kappa values. Therefore, we decided to test whether or not any behavior code was used for a given group. As shown in Table 1, this was a reliable measure (kappa = 0.823). A Chi-square test of independence was calculated comparing the tangible and multi-touch conditions; the tangible condition was more likely to have a behavior code associated with it, $\chi(1) = 6.377$, $p = 0.012$.

VISITOR OBSERVATIONS

In this section, we collate some notable events or themes that were jotted down during visitor interaction sessions in the free-form note taking area of the observation sheet (Figure 6). We witnessed various turn taking strategies with some groups explicitly stating or assigning whose turn it was to use the exhibit. In many cases, the whole group interacted with

Codes	**Multi-touch (count)**	**Multi-touch (%)**	**Tangible (count)**	**Tangible (%)**
*Python**	4	10.26%	4	9.30%
Photo/Video	3	7.69%	9	20.93%
*Dance/Sing**	6	15.38%	9	20.93%
Cards	10	25.64%	17	39.53%
Groups w/ 1 or more codes	16	41.03%	24	55.81%

Table 4: Summary of behavior codes that were recorded for the visitor groups stopping at either exhibit. * indicates unreliable measures.

the system in a given moment. We also saw instances where an individual within a group tried to stake a claim to the exhibit. Somewhat amusingly, one teenager tried to prevent both his father and his brother from "messing up the beat", but after a few minutes of experimentation called for both of them to come back and listen to his creation.

There were other strategies common to both the conditions. For example, after walking up to the exhibit, some visitors would sweep all the balls off the table (in the tangible condition) or deactivate all cells (in the multitouch condition) to start from scratch, while some would start experimenting by adding to the beat left by the previous group. Generally, whenever visitors used beat cards, they also tended to start from scratch. It was also common for children to work together to completely fill the tangible exhibit with balls.

While groups went through a discovery ("how does this work?") phase with both conditions, we noticed instances of problematizing [20], e.g. "I wonder which color is making that sound" followed by a verbal confirmation after figuring it out: "Oh, this is the clap", "and this is the kick".

There were also some interactions patterns unique to the tangible condition. Multiple visitors held a ball up towards the camera or deliberately blocked certain balls from the camera's field of vision. Two teenagers (in two different groups) used this strategy of occluding areas of the exhibit (by waving their arms) to change their beats on the fly; an affordance exclusive to the tangible version of the exhibit. Similarly, the tangible version seemed to prompt more questions about how it worked, not too surprising given the prevalence of multitouch interactions in everyday life.

Visitor interaction vignettes

Group A: 5 children - 3 boys and 2 girls, and a woman interacted with the tangible system for just over 11 minutes. The group demonstrated high levels of engagement and made statements that reflected discovery, collaboration, and excitement. Initially, the children added balls to the existing board. The youngest girl in the group (~8 years old) said, "This is so good"; "I want to fill this whole thing up". Together with two of the boys, the girl proceeded to fill up the whole board. The older girl joined too and said, "This is so cool, I want to keep doing this". A few minutes later, the youngest girl stated that the system reads the color of the balls. The oldest of the children, a boy (~10 years old) then said, "I am like a DJ" while moving the balls around. The woman, one of the boys and one of the girls danced as the exhibit played along. Finally, the woman took several photos of the group around the exhibit.

Group B: A boy and a girl worked together in two phases. First, they played around with the system, tested different ideas and made their own beat. In the second phase, they chose to use the beat cards and continue to make several beats. As they stood next to the tangible exhibit, the boy said that the system is running Python code. The two placed some balls along columns, and eventually filled up the entire board. Once the board was full, they swapped some colors around to alter the sounds (i.e. they used a subtractive rather than additive strategy). They then proceeded to use multiple beat cards and filled out the board accordingly.

Group C: The following vignette includes two women and a man working with the graphical exhibit for over 12 minutes. All three interacted with the screen until a beat emerged. After several minutes one of the women said, "Someone give me a microphone right now". The man added snares to the beat and said, "I am basically coding". One of the women picked up a beat card, and all three collaborated to fill the board based on the beat card. Later the man said "I can do this all day" expressing enjoyment and engagement. Before the group left, they each recorded a video of the exhibit as it played their beat.

There were other sessions where visitor groups became so thoroughly engaged with the system that they turned the tinkering and beat making session into an impromptu musical performance for the whole family (Figure 8).

DISCUSSION

To briefly recap the results, we found that the tangible exhibit had significantly longer engagement times and attracted visitor groups with more children. Visitors engaged in significantly more behaviors indicative of engagement at the tangible exhibit. There were, however, no significant differences in the portion of active visitors in groups between the two exhibits. This suggests that the tabletop display and

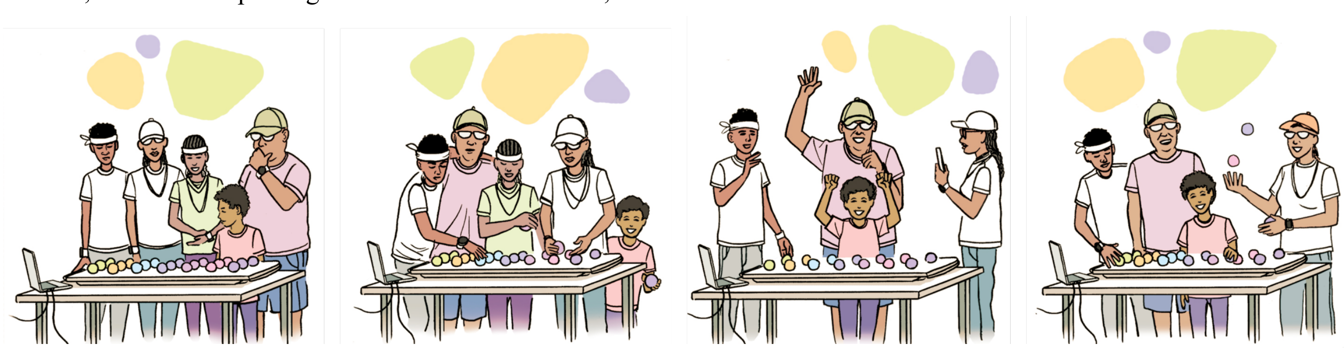


Figure 8: A family figured out the exhibit, danced along to the beat they came up with. The father also freestyle rapped to the beat while another family member recorded this performance on their mobile phone (sketches derived from video recordings to *preserve anonymity*).

tangible exhibit were equally effective at supporting multiple simultaneous users.

Horn [12] suggests that tangible objects can be particularly effective at evoking *cultural forms* of literacy, learning, and play. These cultural forms in term cue physical, social, and emotional resources that visitors use to make sense of exhibits and structure the shared activity. Based on this theory, we could interpret our results as the physical plastic balls being more effective at communicating the idea of a fun and playful activity than the corresponding touchscreen display. Our observation data also suggested that there were a number of playful ways that visitors could physically interact with the balls (including juggling, sorting, filling up all of the slots on the table, or even just taking the balls on and off the table for younger children). These diverse and parallel modes of engagement likely contributed to keeping family groups engaged together. It is important to note that a major reason visitors leave an exhibit is because they are pulled away by another family member. The tabletop display was, however, equally effective at involving multiple visitors in the same group suggesting that such technologies can be useful for supporting collaborative and joint engagement. And, from our observations, visitors seemed to enjoy the exhibit regardless of the condition. We also noticed a degree of spectatorship and performance in both conditions. It was almost as if visitors were playing a musical instrument for an audience.

We close with a note about simplicity in design, especially for museums. Our extremely minimal approach left out many features such as tempo adjustment, audio filters and effects, or even numbers to indicate beat counts. Despite this, the exhibits were surprisingly compelling for visitors. In short, the design was simple, but not simplistic.

CONCLUSION

We presented a study comparing the use of a tangible and a multitouch interface in a museum exhibit designed to explore musical rhythms. Both the exhibits were designed while adhering to a principle of minimalism, and they led to an engaging visitor experience. We found that the tangible exhibit had significantly longer engagement times and attracted visitor groups with more children. Also, visitors engaged in significantly more behaviors indicative of engagement at the tangible exhibit. These results corroborate and build upon Horn et al's [14] research comparing TUIs to a single-mouse based digital interface. In this case, both the tabletop display and tangible exhibit were equally effective at supporting multiple simultaneous users.

LIMITATIONS AND FUTURE WORK

While quantitative analysis was a primary focus of this research, we think that in-depth qualitative research based on interviewing museum visitors or doing video analysis can yield other design recommendations. Our results were collected in a museum that usually gets more than 1.5 million annual visitors from diverse backgrounds; however, the collected data is from one museum, in one city in the US. There might be limitations on generalizing the results to other contexts.

Our prototypes were on two extremes of the TUI and GUI spectrum. We think that exploring design options that blend between these modalities would be worth researching in the future.

ACKNOWLEDGEMENTS

The authors would like to thank Brian Magerko and Jason Freeman from Georgia Institute of Technology, and Aaron Price from the Museum of Science and Industry, for their assistance with interface design and data collection. This research was supported by grants DRL-1612619 and DRL-1451762 from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

SELECTION AND PARTICIPATION OF CHILDREN

We tested our exhibit at a science museum and this exhibit was open to all museum visitors, including families and children. The children who used the system were self-selecting or accompanying friends and family.

REFERENCES

[1] Shiroq Al-Megren and Roy A. Ruddle. 2016. Comparing Tangible and Multi-touch Interaction for Interactive Data Visualization Tasks. In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16). Association for Computing Machinery, New York, NY, USA, 279–286.

[2] Alissa N. Antle and Sijie Wang. 2013. Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. In Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction (TEI '13). Association for Computing Machinery, New York, NY, USA, 65–72.

[3] Rene-Piere Bardet. 1986. 200 Drum Machine Patterns, Hal-Leonard Publications.

[4] Florian Block, James Hammerman, Michael Horn, Amy Spiegel, Jonathan Christiansen, Brenda Phillips, Judy Diamond, E. Margaret Evans, and Chia Shen. 2015. Fluid Grouping: Quantifying Group Engagement around Interactive Tabletop Exhibits in the Wild. In Proceedings of Conference on Human Factors in Computing Systems (CHI '15). ACM, 867-876.

[5] Kevin Crowley, M. Callanan, M, Jipson, J. L., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. Science Education, 85(6), 712-732.

[6] Catherine Eberbach and Kevin Crowley (2005). From living to virtual: Learning from museum objects. Curator: The museum journal, 48(3), 317-338.

[7] Randi A. Engle & Faith R. Conant. 2002.Guiding Principles for Fostering Productive Disciplinary Engagement: Explaining an Emergent Argument in a Community of Learners Classroom, Cognition and Instruction, 20:4, 399-483.

[8] George W. Fitzmaurice and William Buxton. 1997. An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97). Association for Computing Machinery, New York, NY, USA, 43–50.

[9] Amanda Harris, Jochen Rick, Victoria Bonnett, Nicola Yuill, Rowanne Fleck, Paul Marshall, and Yvonne Rogers. 2009. Around the table: are multiple-touch surfaces better than single-touch for children's collaborative interactions? In Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1 (CSCL'09). International Society of the Learning Sciences, 335–344.

[10] Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale. EMDialog: Bringing information visualization into the museum. IEEE transactions on visualization and computer graphics 14, no. 6 (2008): 1181-1188.

[11] Hinrichs, Uta, and Sheelagh Carpendale. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3023-3032. 2011.

[12] Michael Horn. 2018. Tangible Interaction and Cultural Forms: Supporting Learning in Informal Environments, Journal of the Learning Sciences, 27:4, 632-665.

[13] Michael Horn, Zeina Atrash Leong, Florian Block, Judy Diamond, E. Margaret Evans, Brenda Phillips, and Chia Shen. 2012. Of BATs and APEs: an interactive tabletop game for natural history museums. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 2059-2068.

[14] Michael S. Horn, Erin Treacy Solovey, R. Jordan Crouser, and Robert J.K. Jacob. 2009. Comparing the use of tangible and graphical programming languages for informal science education. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). ACM, 975-984.

[15] Eva Hornecker. "I don't understand it either, but it is cool"-visitor interactions with a multi-touch table in a museum.In 2008 3rd IEEE international workshop on horizontal interactive human computer systems, pp. 113-120. IEEE, 2008.

[16] Thomas Humphrey and Joshua P. Gutwill. Fostering active prolonged engagement: The art of creating APE exhibits. Routledge, 2017.

[17] Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi, and Toni Laitinen. 2010. Worlds of information: designing for engagement at a public multi-touch display. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). Association for Computing Machinery, New York, NY, USA, 2267–2276.

[18] 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. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 201-210.

[19] Zeina Atrash Leong, Michael S. Horn, Lisa Thaniel, and Emily Meier. 2018. Inspiring AWE: Transforming Clinic Waiting Rooms into Informal Learning Environments with Active Waiting Education. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 98, 1–12.

[20] Leilah Lyons, Michael Tissenbaum, Matthew Berland, Rebecca Eydt, Lauren Wielgus, and Adam Mechtley. 2015. Designing visible engineering: supporting tinkering performances in museums. In Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15). ACM, 49-58.

[21] Joyce Ma, Lisa Sindorf, Isaac Liao, and Jennifer Frazier. 2015. Using a Tangible Versus a Multi-touch Graphical User Interface to Support Data Exploration at a Museum Exhibit. In Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15), ACM, 33-40.

[22] David Merrill, Emily Sun, and Jeevan Kalanithi. 2012. Sifteo cubes. In CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12). ACM, New York, NY, USA, 1015-1018.

[23] Johanna Okerlund, Evan Segreto, Casey Grote, Lauren Westendorf, Anja Scholze, Romie Littrell, and Orit Shaer. 2016. SynFlo: A Tangible Museum Exhibit for Exploring Bio-Design. In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16). Association for Computing Machinery, New York, NY, USA, 141–149.

[24] Frank Oppenheimer. 1976. Everyone Is You... Or Me. Technology Review Volume 78,

[25] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. 2008. It's Mine, Don't Touch! interactions at a large multi-touch display in a city centre. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08).

Association for Computing Machinery, New York, NY, USA, 1285–1294.

[26] Daniela Petrelli and Sinead O'Brien. 2018. Phone vs. Tangible in Museums: A Comparative Study. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM.

[27] Sara Price, Yvonne Rogers, Michael Scaife, Danae Stanton, and Helen Neale. Using ‘tangibles’ to promote novel forms of playful learning. Interacting with computers 15, no. 2 (2003): 169-185.

[28] Sara Price and Yvonne Rogers. 2004. Let’s get physical: The learning benefits of interacting in digitally-augmented physical spaces, Journal of Computers and Education, vol. 15, no. 2, pp. 169–185.

[29] Mitchel Resnick, Amy Bruckman, and Fred Martin. 1996. Pianos not stereos: creating computational construction kits. interactions 3, 5 (1996), 40-50.

[30] Lotte de Reus, Jouke Verlinden, Maaike Roozenburg. 2013. Nonlinear stories told by cups and saucers: Smart Replicas with responsive 3D audio. Ar[t] Magazine, 3, 34-39.

[31] Jessica Roberts, Amartya Banerjee, Annette Hong, Steven McGee, Michael Horn, and Matt Matcuk. 2018. Digital Exhibit Labels in Museums: Promoting Visitor Engagement with Cultural Artifacts. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM.

[32] Orit Shaer and Eva Hornecker. 2010. "Tangible User Interfaces: Past, Present, and Future Directions", Foundations and Trends® in Human–Computer Interaction: Vol. 3: No. 1–2, pp 4-137.

[33] Scott S Snibbe and Hayes S Raffle. 2009. Social immersive media: pursuing best practices for multi-user interactive camera/projector exhibits. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 1447–1456.

[34] Philip Tuddenham, David Kirk, and Shahram Izadi. 2010. Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’10). Association for Computing Machinery, New York, NY, USA, 2223–2232.

[35] TrackingJS Computer Vision Library, https://trackingjs.com, last accessed Dec 09, 2019.

[36] Brygg Ullmer and Hiroshi Ishii. 2000. Emerging frameworks for tangible user interfaces. IBM Syst. J. 39, 3-4 (July 2000), 915-931.

[37] Lesley Xie, Alissa N. Antle, and Nima Motamedi. 2008. Are tangibles more fun? comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In Proceedings of the 2nd international conference on Tangible and embedded interaction (TEI '08). ACM, 191-198.