Combining Web, Mobile Phones and Public Displays in Large-Scale: Manhattan Story Mashup

Ville H. Tuulos[1], Jürgen Scheible[2], and Heli Nyholm[3]

[1] Helsinki Institute for Information Technology tuulos@cs.helsinki.fi
[2] University of Art and Design Helsinki jscheib@uiah.fi
[3] Nokia Research Center heli.nyholm@nokia.com

Abstract. We present a large-scale pervasive game called Manhattan Story Mashup that combines the Web, camera phones, and a large public display. The game introduces a new form of interactive storytelling which lets an unlimited number of players author stories in the Web while a large number of players illustrate the stories with camera phones. This paper presents the first deployment of the game and a detailed analysis of its quantitative and qualitative results. We present details on the game implementation and set up including practical lessons learnt about this large-scale experiment. The analysis shows how the game succeeds in fostering players' creativity by exploiting ambiguity and how the players were engaged in a fast-paced competition which resulted in 115 stories and 3142 photos in 1.5 hours.

1 Introduction

Manhattan Story Mashup (MSM) combines the web, mobile phones and one of the world's largest public displays in Times Square to a large-scale pervasive game. The game was played by over 150 players in the web and 184 streets players in Midtown Manhattan. The web players used the storytelling tool at the game web site to mash up stories, either by writing new sentences or by re-using already illustrated sentences. A noun from each new sentence was sent to a street player for illustration. The street player had to shoot a photo which depicted the word within 90 seconds. The photo was then sent to two other street players who had two guess what the photo depicts amongst four nouns, including the correct one. If the photo-noun pair was guessed correctly, the original sentence was illustrated with the new photo and it turned into an ingredient for new stories. The best stories were shown on the Reuters Sign in Times Square in real-time.

The game introduces a new form of interactive storytelling which lets distant people to collaborate in real-time. The web players get a real-time human-mediated sneak peek to the physical world which they may steer at a desired theme. The street players, who are taking the photos, may use their imagination at the fullest while trying to find the requested targets in a fast-paced competition. Since both the requested individual nouns and the returned photos are

highly ambiguous in nature, the game feels somewhat mysterious, yet meaningful and exciting to all the players. On the web player's viewpoint, this may be seen as *collaborative leisure* [1] whereas the street player may see it as an urban game.

Manhattan Story Mashup brings together many concepts from earlier pervasive games. It links the physical and virtual worlds, as the games described in [2], and engages the players in a collaborative and competitive effort of storytelling [3]. It also entices players to share their experiences through a public display [4, 5]. and provides an entertaining and motivating context to produce experimental data for further purposes [6].

MSM is a part of the SensorPlanet project at Nokia Research Center. In its origins, it was motivated by the need to understand issues related to mobile phone-centric sensing. Especially we wanted to get hands on vast amounts of real-world data, collected by actual mobile phones. Design, implementation and orchestration of the game provided valuable knowledge on experimentation with a pervasive application in the real world.

The game produced a rich set of data. All game events were collected in a database which allowed us to analyze the whole game event afterwards in detail. Immediately after the game, the street players were asked to fill in a questionnaire which contained questions about the game design and experience. Analysis of both the quantitative and qualitative data is presented in section 5. Some of the design decisions proved to be remarkably successful; they are listed in the conclusions. We also outline the system implementation and the set up process which present some practical lessons learnt.

We hope that the scale and the multi-faceted nature of Manhattan Story Mashup provides useful information for the design and orchestration of future experiments which utilize mobile phones, web and / or public displays. Central contribution of this paper is to strengthen some earlier findings, such as benefits of ambiguity, and to present a new field-tested game design which provides a working example of collaborative leisure and co-presence.

2 Prior Work

Positioning has been a defining feature in many earlier pervasive games. For instance, Pirates! [8] uses the physical world as a game board in which the game takes place. In this constrained game arena locations of the players may be determined using short range radio beacons. Yoshi [9] and Bill [10] exploit spotty coverage of WiFi networks in a clever way by taking advantage of this seeming limitation. In these games, the gameplay is designed around the concept of physical location. MSM was about to utilize locationing as well but only in a minor role. However we could not come up with a technically straightforward locationing method which would have fitted in the game without taking too much attention from the key concepts. Despite of this, we logged internally the GSM cell IDs of the players during the game.

MSM is first and foremost an urban photo hunt with a twist. A similar concept, involving both camera phones and public displays, was sketched by PhotoPhone Environment [11], although they did not present any implementation. MSM and Snagu[4] share the common concept of "reverse-Flickr": Given a keyword or tag, the player takes a photo resembling the word. Sharing the Square [1] is a system for sharing leisure between distant people through photographs, voice and location. The MSM concept is tangential to the Square's idea of implementing co-presence through an interactive photographing process. Their treatise of shared photographing is relevant for MSM as well.

Manhattan Story Mashup builds on previous works of research which combine the physical and the virtual into a seamless game experience. Two large-scale games were produced in collaboration between the artist group Blast Theory and the Mixed Reality Laboratory at the University of Nottingham. In Can You See Me Now [12] online players are chased through a virtual model by street players who play in a real urban environment. Similarly, Uncle Roy All Around You [7] involves players both in the field and in a parallel virtual world. Street players were given a task to search for a character named Uncle Roy. Remote players, together with professional performers, guided the street players in their quest. Experiences from these games provide a solid background for designing and orchestrating collaboration in MSM.

Both the previous games are carefully orchestrated and controlled together with professional performers, which play a major role in these games. In contrast, our motivation was to gain understanding in spontaneous behavior of players, both in the Web and in the streets. Therefore we deliberately left room for emergent features and unexpected events to happen. Furthermore we designed MSM so that there would not be any inherent limitations in number of participating players. This sets MSM apart from many previous games; for example Can You See Me Now can be played only by fifteen online players at time. We hypothetize that this approach might give us realistic data on real-world sensing and user behavior, which may be generalizable to non-game related settings as well. The approach is akin to some previous games for collecting "serious" data, such as the ESP game for labelling images in the web [6].

As an example of interactive storytelling on mobile devices, Bamford et al. [3] reported an innovative mobile game in the form a multi-authored mobile book based on the 1920s surrealist technique of *Exquisite Corpse*. The book builds from a series of standard text message length contributions, each author being given only the previous message on which to base their own contribution. MSM goes a step further by including the images illustrated by some people and stories written or mashed up by other people, using recent techniques of the social web.

The usage of public displays haven been researched in various contexts. One of the main challenges associated with interactive public displays is how to entice people to interact with them [4]. Another challenge is how to share a single public display between multiple users. This is where personal mobile phones come handy in terms of dispersing access and control. For example, Scheible

[4] http://snagu.com

and Ojala [5] demonstrated with MobiLenin a solution for realizing multi-user interaction with a public display using personal mobile phones.

3 Game Design

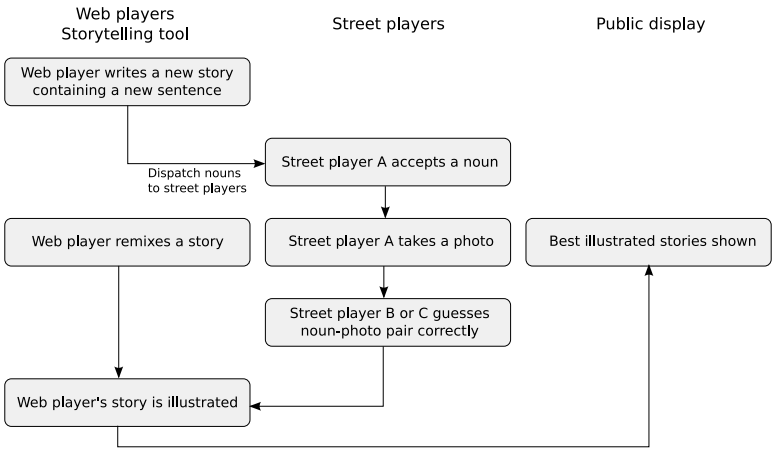

Fig. 1. Gameplay

The overall goal from the perspective of a street player in MSM is to collect points by shooting photos for illustrating stories written by people in the web. The player who collects the most points is the winner of the game. There are two tasks for the street player: One task is to shoot photos representing nouns that are extracted from story sentences created by people on the game's website in real-time. The mobile phone receives these nouns automatically and shows them as keywords in a list on the screen and the player just needs to select a keyword. The phone camera will then open automatically and start a countdown timer that grants between 60-90 seconds time to shoot the photo. The photo is then sent automatically to the game server. The other task is to match keywords to images taken by other players to validate the quality of the images going into the system. The screen will show an image and a list of four keywords including the correct one. The player needs to guess which of the keywords matches the image. The image is sent at the same time to another player for guessing in order to let the players compete. The first player to guess the word-photo pair correctly gets the points.

The player can gain different amounts of points: 1 point for shooting a photo, 6 points for guessing correctly, 9 points for a photo that is taken by the player himself and correctly guessed by some other player. The points are accumulated during the whole game. The players were able to see their own score and rank all the time on the lower part of the phone screen.

The core of the game is to produce imaginative noun-photo pairs. The rationale for guessing is to make sure that the photos actually represent the desired target, or that the association between word and photo is conceivable by another human being, even though it may be highly ambiguous. The reason for having two players to guess the same photo is simply to increase likelihood of having a successful guess and to motivate each player to guess well – otherwise a competitor would gain more points.

The gameplay is illustrated in Fig. 1. If a storyteller in the web chooses to use only already illustrated sentences by other people, no street player actions are needed. If the storyteller contributes new sentences, a noun from each sentence must go through the illustration process. If the street players fail to accomplish any step, the story will not finish since at least one sentence will be left unillustrated.

In case that the web players are unable to keep the street players busy by writing enough stories, there is a backup mechanism which dispatches nouns to the street players from a predefined list. If such an automatically dispatched noun passes the illustration process, it may be used in a new sentence without additional delay.

3.1 Discussion

Based on their experiences with Uncle Roy All Around You, Benford et al present a design framework for mobile experiences [7]. In the following, we reflect main points of our design to this framework.

We wanted to avoid the situation in which the web players would be mere spectators, while the street players would be the *de facto* performers. Games like Uncle Roy and Can You See Me Now provide the remote player a virtual model of the physical city, thus mimicking the physical experience to some degree. In contrast to this approach, our design provides two different, yet equally important facets to the game, both respecting the natural context of action: The web player feels that she is participating in a Web 2.0-ish collaborative effort, so she may well regard herself as a major performer in the game. In parallel, the street player takes part in the hectic urban game which makes her another true performer. This two-sided approach is akin to so called *seamful design* [10] of ubiquitous systems. Instead of trying to hide differences between the virtual and physical worlds, we try to exploit the best features of both worlds.

However, both players are also spectators. Following the taxonomy presented in [13], we consider the street player to have a *suspenseful* view to the game, since she is unable to see the effects of her actions immediately: She takes a photo and hopes that someone guesses it correctly and it will get integrated to a story. In the web player's point of view, the game is *magical*, since she cannot identify the exact source of the photos: She sees newly illustrated sentences, written by other web players popping up every now and then and her own sentences becoming illustrated by some random street players.

The large public display supports the suspenseful and magical nature of the game. After some delay the street player may see her own images on the display

as a part of a story. The web players could see their stories through the Reuters' webcam magically presented in a physical place in some distant location.

Another important feature in our design is deliberate ambiguity in the tasks. This approach is suggested for game design in [7] and [14]. Instead of forcing or encouraging the web players to use only unambiguous and concrete words, such as "house", "milk" or "sun", we picked a random noun from each sentence. This made the game more exciting and left plenty of room for players' creativity. Ambiguity is especially apparent in the guessing part of the game which requires a human player to interpret photos: Consider an image of a building. Is it a hospital or a dormitory? All these tasks boil down to being able to guess another player's intent, even though the message is mediated through an ambiguous channel, namely through a word or a photo [14]. It is even desirable that some ambiguity remains in the resulting stories since the results are often hilarious.

4 Implementation

4.1 Storytelling Tool

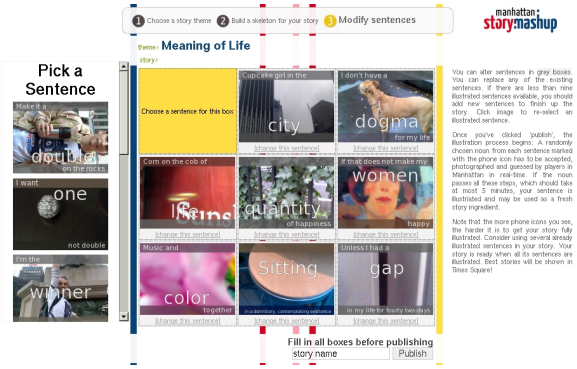

Fig. 2. Storytelling tool in the web

The Storytelling tool lets an individual web site visitor to leave her creative handprint to a large public sign, as well as interact with the real people in the streets of Manhattan. However presenting just a collection of individual contributions would be rather uninteresting. Also the required effort to make an individual contribution from the scratch might be prohibitively high for many. Instead, the storytelling tool lets the user pick a previously contributed story and use it as a basis for her own story. Alternatively one can use previous sentences to mash up or remix a personal story.

To enhance usability and approachability, the first steps with the tool were made as convenient as possible. No registration was needed: There was a link in

the front page leading directly to the tool. There were no options from which to choose. The user could start playing around with the tool immediately and she could contribute her own sentences with no restrictions.

A story consists of nine illustrated sentences as shown in Fig. 2. For each cell or slot the user may either pick an already illustrated sentence, written by someone else, or she may write a new sentence. Once the user has chosen appropriate content for each slot and the resulting story looks somewhat meaningful, she may publish the story.

It is practically impossible to detect dubious or nonsensical sentences automatically as the problem is highly semantic in nature. However, we performed some simple filtering using a blacklist of 333 common profanities. We also made some syntactical checks, ensuring that the sentences are long enough and do not contain any meaningless characters.

After publishing, each new sentence is tokenized. Each token is checked against WordNet [15] and nouns are collected. From each new sentence a random noun is chosen for illustration. The nouns, the number of which may vary from zero to nine, are dispatched to random players in Manhattan, yet taking care that no player has more than ten nouns at time. After this the story moves to the pool of incomplete stories, until all its sentences are illustrated. For an overall view to the gameplay, see Fig. 1.

Storytelling tool has an important feature that should keep the stories meaningful, even though users are free to write anything to sentences: When the user picks sentences for her story, she is presented a list of available illustrated sentences to choose from. This list is ordered by descending popularity: The more frequently a sentence gets picked, the higher ranking it will have in the list. This feature is similar to filtering mechanisms found in many social bookmarking sites, such as Digg[5] or Reddit[6]. The rationale is that even though someone may find entertaining to input nonsense to the system, almost no one regards nonsense written by an unknown person interesting. Thus nonsense gets disregarded by many and its ranking drops. This phenomenon is further amplified by the fact that most users consider only the top entries in the list, being too lazy to browse everything through, and thus increase the popularity of the already popular items.

4.2 Mobile Client

We decided to use Nokia N80 mobile phones in the game, mainly due to their new S60 3rd edition software platform, WiFi support and high-quality 3 megapixel camera. The game client software was implemented in Python for S60 (PyS60) which is an open-source port of the Python programming language to the S60 platform [16]. Python was chosen due to its suitability to rapid prototyping and easy extendibility in native C++. As described below, source code availability was an important factor as well.

[5] http://digg.com
[6] http://reddit.com

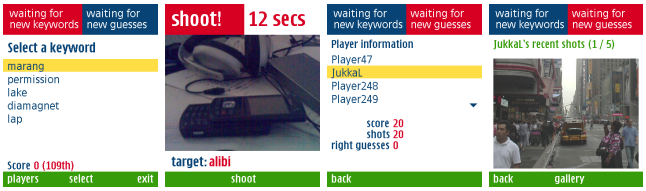

Fig. 3. Mobile Client: Keyword selection mode, shooting mode, player statistics and gallery views

We implemented a set of custom UI widgets for the game instead of using the standard UI library. This gave us more flexibility in usability design and made the client look and feel more game-like. The standard PyS60 distribution includes support for taking photos, but does not provide the viewfinder. Since smooth camera usage is central to the game experience, we implemented viewfinder support as an extension. Thus the player did not have to use any other phone functionality or software besides our game client, which greatly enhanced game immersiveness.

The game client has three modes: Keyword selection mode, shooting mode and guessing mode. The game leads the player from one mode to another. In the keyword selection mode player chooses one of the available nouns as the next target for shooting. The keyword choice activates the shooting mode which opens the viewfinder and lets the player to find a suitable target. Once the photo is taken, it is sent automatically to the server and the player returns to the keyword selection mode.

The game client polls the server every five seconds to update the list of keywords and retrieves requests for guessing. Once a new request for guessing is received, the guessing mode is activated automatically unless the player is shooting photo. In the guessing mode the player is presented four alternative nouns, one of which is the correct one, together with a photo taken by another player. After choosing one of the alternatives, the player is taken back to the keyword selection mode.

In each of the modes, there is a visible countdown timer which forces the player to make quick decisions. In the selection mode, the timer expires keywords which have been shown for over 90 seconds. In the guessing and shooting modes, the player is taken back to the selection mode if she was unable to act in 60 seconds. In technical point of view, tight timeouts make sure that players get a constant stream of fresh tasks and passive players cannot stall the game dynamics. Rather surprisingly, since players see that the game keeps going without their explicit action, they feel motivated or even urged to act. This was a major factor in making the game highly engaging.

In addition to the three main modes, there is a screen showing current player statistics. A simple gallery is provided so that the player could see the most recent photos taken by other players. These features were added to increase feeling of

competitiveness and collaborative effort. In practice however, the single line in the keyword selection mode showing the player's current score and ranking in real-time proved to be sufficient for this purpose.

An important detail in large-scale field experiments involving mobile phones is how to set up a large number of devices. Installing software to 200 phones manually is not impossible, but it is a remarkable feat. There are some standardized methods for large-scale software deployment for mobile handsets, such as Over-The-Air Programming, but often these methods are only available for operators and phone manufacturers.

Our approach was twofold: First, we were able to compile a customized version of Python for S60 since its source code is freely available. We modified the user interface of the interpreter to include functions for game client update and launch. Also the viewfinder extension was included in the new build. This reduced the number of packages which needed to be installed. Secondly, we were able to automatize the installation process to some degree by using freely available Obexftp[7] tool for Linux to transfer the installation packages to phones. We considered using several Bluetooth dongles to transfer packages to multiple phones simultaneously, but this proved to be somewhat unreliable in Linux. Instead, we used a USB-hub to connect five phones to a laptop at the same time, which greatly reduced time needed to transfer the files. After one has figured out the process, we can estimate that installing packages to some 200 phones would take 10-15 person hours. However other tasks, such as sorting out the SIM cards and recharging the phones, took two days from three persons.

4.3 Large Public Display

We rented the Reuters Sign[8] in Times Square for exclusive use during the game. The display system consists of 11 individual displays which may be used either separately or as a single large display. This display was chosen due to its prominent location and enormous visibility. Times Square is an iconic location in global scale, thus the possibility to create personal content to be shown there was attractive for people around the world. Considering the storytelling tool, it was crucial that the web users could relate to the place where the stories were to be shown, even though the location was distant.

Once all the nine sentences of a story had successfully gone through the illustration process, depicted in Fig. 1, the story became a candidate for presentation on the display. We had a human moderator filtering the candidate stories using a separate moderation interface in the web. The interface allowed the moderator to "bless" a story for presentation, or to blacklist individual sentences containing unwanted content.

Once a story was blessed, it was automatically sent to the display system by the game server. Depending on the display status, the new story was shown there after some 1-5 minute delay. All the nine illustrated sentences were shown

[7] http://openobex.triq.net
[8] http://www.timessquare2.com

at once on the display. One by one, each sentence was enlarged and shown on the large middle display for six seconds, as seen in Fig. 4.

The display added a unique twist and a big wow-effect to the game. It provided a feedback channel for the street players who were able to follow in real-time how their photos were intervowen into various stories. This role, providing a shared view to the game, was probably the most important feature of the display. In addition, the unique opportunity of getting a personal fingerprint to Times Square motivated both the web and the street players to produce imaginative content.

Fig. 4. The Reuters Sign in Times Square. Photos by Kitty Kat, Jürgen Scheible and _snipp.

5 Experience

The actual game event took place on September 23rd 2006 between noon and 1:30pm in Midtown Manhattan. The game was one of the featured games in Come Out and Play street games festival [9]. We had invited 140 university students to participate from New York University, Parsons The New School for Design and Brooklyn Polytechnic. In addition we had invited some 70 persons from various companies and institutions to join the game.

In total 184 players played the game. During the game the players took 3142 photos and made 4529 guesses, 2194 (48.4%) of which were correct. Technically, there should have been 6284 guesses in total, assuming that every photo was guessed by two separate players. However, if all the players were busy, i.e. taking a photo or already guessing, only one guesser would suffice. In the extreme case that no guessers were available, the image was accepted as such. Also if the guessing timeouted or the player closed the client without guessing, no guess was recorded.

Figure 5 visualizes the whole gameplay. Each row in the graph corresponds to a player. Rows are ordered by descending score, thus the topmost row corresponds to the winner of the game. X-axis is time, from the game start to the end,

[9] http://comeoutandplay.org

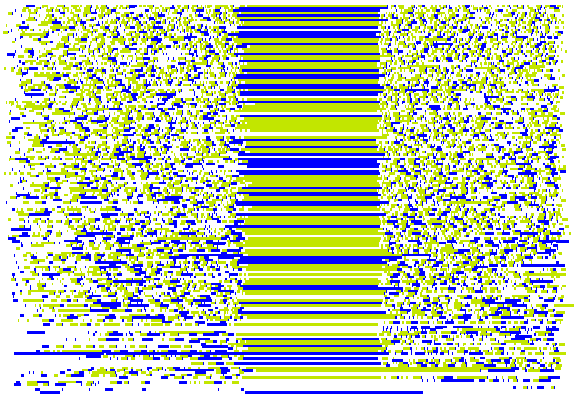

Fig. 5. Gameplay visualized: Rows correspond to players, x-axis is time. Green/light segments denote time spent taking photo and blue/dark segments denote guessing. Rows are ordered by descending score.

4902 seconds or 81 minutes in total. Color segments indicate player action at every moment, which correspond to the three modes of the game client: White indicates keyword selection mode, green/light shooting and blue/dark guessing.

The most distinctive feature in the graph is the large gap in the middle. This corresponds to an unknown bug which appeared after the game had been running for some 30 minutes. Game server stopped dispatching nouns, apparently assuming that all players already have the maximum amount of pending keywords. It took a while to notice that the bug affected all the players. After this, the game server was restarted and the game history was re-evaluated, which normalized the situation and the game continued.

The following phenomena can be interpreted from the graph:

1. Gameplay was fair: All the players seem to have done comparable amount of shooting (green/light) and guessing (blue/dark).
2. Nouns were not too hard: There are only few long white gaps in player actions, thus players accepted proposed nouns quickly.
3. Players were motivated: Even the bottommost players played hard all the time.
4. Players did not get bored: Right-hand side of the graph looks the same as the left.
5. Scoring is not random: Segments are slightly longer in the bottom part of the graph, indicating that low scores correlate with slow actions.

Observation 1. proves that the game dynamics worked correctly: It was a deliberate decision that a player could not avoid guessing by staying in the shooting mode all the time. Observation 2. proves a hypothesis correct: Players were remarkably creative with the words and they were able to illustrate even the most abstract words, such as “synergy”, by utilizing the rich urban environment and

spontaneous acting. Observation 3. was unexpected: We assumed that the players who notice their ranking to be rather low, say 83th, would soon lose interest in playing. However it seems that the players often got spurts of easy targets, leading to sudden increase in ranking which kept motivation high. This leads to the conclusion that showing the current ranking in the game UI is a good idea. Similarly observation 4. was unexpected: We decided to make the game short enough, so that people would not get bored during the game. Clearly the game was engaging enough to keep players fully focused for 90 minutes. Observation 5. was expected, but we assumed that the effect would be much stronger. The data shows that the players were surprisingly homogeneous, regardless of their final score.

We analyzed the game history also quantitatively, to confirm the observations from the graph and to find hidden patterns. Basic statistics over players are presented in Table 1. Player characteristics were distributed rather uniformly and no clusters of different types of players were found, which agrees with observation 3. Overall, it seems that that the players were a homogeneous group of well-motivated individuals which resulted to a uniform outcome. This provides us a solid baseline of the game design against which different, more heterogeneous groups of players may be evaluated in the future.

We were interested in finding out whether the winning players had some special characteristics compared to the others. We found no clear pattern related to number of correct guesses or guessing in general. The most remarkable correlation exists, not surprisingly, between the average shooting time and the final score. To conclude, it seems that the winners won by making good decisions quickly and by acting fast. This might be considered as a positive outcome for a new game design.

Table 1. Gameplay statistics over players

Variable	Min	Max	Median
Nouns chosen	2	64	35
Photos shot	2	58	32
Guesses	3	56	28
Correct guesses	1	29	14
Score	18	430	246
Guessing time (avr secs.)	6.5	39.3	15.0
Shooting time (avr secs.)	15.9	91.9	33.4
GSM cells visited	1	26	14

We also recorded player locations based on GSM Cell IDs. Even though the game took place in a restricted area, namely between the 59th and 43rd streets, around the 5th Avenue and Broadway, in total 197 unique cell IDs were recorded. There seems to exist a weak positive correlation between the number of cells visited and the final score.

5.1 Photos

The most intriguing tangible result of the game were the 3142 noun-photo pairs. The game was designed to stimulate creativity. Our research motivation for this was to get some preliminary ideas on what kind of "action possibilities" urban environments might provide and what kind of non-trivial features people are able to find in their surroundings.

We analyzed 523 photos manually, to gain understanding in the types of associations present in the photos. The largest single theme was 193 photos showing human beings, or players in our case. Sometimes acting was used to set out a role, facial expression or a character, as in the leftmost photo in Fig. 6. However, the photos did not always depict a human being as such, but another player acting out an abstract concept such as "evidence", "link" or "equal" as seen in the middle photo. Sometimes the association was related to the textual form of the word, resembling a wordplay, as in the rightmost photo.

We were surprised to see how effortlessly the players were able to cross the boundaries of ordinary public behavior, e.g. by acting out publicly, once they started to look at the world through the game client, especially when playing in a team. This suggests that concerns related to blurring the line between the virtual and physical, which are discussed e.g. in [7], are valid even though the game by itself may be quite harmless and abstract, as in our case.

Fig. 6. Example photos and sentences

Another imaginative approach to deal with abstract concepts was to deliberately blur or shake the photo to hide irrelevant details. Of the analysed photos, 155 were shaky. Shaky photos were also used to depict colors or movement, as in "explosion" and "speed" or kicking something to show "temper". In case of unambiguous but unavailable objects, such as "mustard", "film" or "balloon", players tried to blur another object with some resemblance to the target object to produce the desired effect. Naturally some photos were shaken or out of focus by accident. This happened often with close ups ("powder", "ear", "bruise") if the player forgot to switch the macro mode on. Since the players were forbidden to use any visible trademarks or copyrighted items in the photos, some players tried to blur the offending target to circumvent the rule.

Since the nouns were recognized using WordNet, which also includes semantic relations between words, we were able to group nouns under some generic categories or themes automatically. To reduce the number of categories shown, Table 2 presents only those categories or hypernyms which were used to extract words for automatic dispatching before the game. The numbers in parentheses indicate the total amount of nouns illustrated in the corresponding category. Accuracy indicates the percentage of correctly guessed noun-photo pairs in the category.

It is worth noting that the lowest accuracy is well above the default, 0.44, which is the probability that at least one of the two guessers makes the correct choice assuming that the choices are random. Since the actual accuracies are above this, we can assume that the players were paying attention to the guessing part and probably tried to perform as well as possible.

Table 2 shows that explicit human-related subjects, such as facial expressions or characters, are easy to act out and photograph regardless of the surroundings. Likewise unambiguous concrete objects, such as bodies of water, shops, and beverages are easy to guess. In contrast, objects that are difficult to set out, such as "dormitory" or "breakfast", produce often ambiguous photos. However, the players seemed to enjoy ambiguity and in some cases they deliberately took ambiguous photos for the sake of fun and ingenuity. Thus although ambiguous words and photos made the game more difficult, they were likely to provoke engagement and exploration in the game [13].

Table 2. Guessing accuracies per WordNet categories. Total number of guesses in parentheses.

Hypernym	Accuracy	Hypernym	Accuracy	Hypernym	Accuracy
facial expression	0.94 (17)	car	0.74 (34)	currency	0.67 (3)
character	0.86 (7)	show	0.72 (25)	friend	0.66 (35)
writing implement	0.85 (13)	bread	0.71 (14)	edible fruit	0.64 (28)
body of water	0.84 (31)	wheeled vehicle	0.69 (36)	light	0.64 (67)
athlete	0.83 (24)	waste	0.69 (29)	road	0.62 (29)
shop	0.81 (26)	motor vehicle	0.68 (19)	toy	0.61 (18)
performer	0.77 (22)	tool	0.68 (50)	material	0.61 (79)
chromatic color	0.77 (61)	human	0.68 (499)	piece of work	0.59 (22)
beverage	0.75 (24)	garment	0.68 (87)	dish	0.56 (48)
figure	0.74 (72)	device	0.67 (181)	building	0.56 (70)

5.2 Web

Approximately 4000 unique IP addresses had visited the game web site at storymashup.org before the game launch. The storytelling tool was open only during the game for 90 minutes. During this time 165 unique IP addresses visited the

game web site. 115 stories were published, which were mashed up using the 271 sentences written during the game.

Since the players in Manhattan accepted 5603 nouns in total, only 4% of the words originated from the new sentences. However, the storytelling tool offered a possibility to use already illustrated nouns in a sentence. This feature was used 129 times. We hand-picked 26 best stories which were shown in Times Square. The game design, as well as the implementation, would have been able to handle much larger number of storywriters in the Web.

A major technical research motivation for having the web site in the first place was to gain better understanding in different time-scales of the web and the physical world, and how it affects the system dynamics. In practical terms, minutes make a big difference while one is standing on a busy street in Manhattan, compared to web site which is still mostly conceived as a rather static entity. Especially it is not accustomary to have a web site open only for 90 minutes, even though such a happening makes sense in the physical world. Thus to achieve smooth real-time interaction between the web and the physical world, the system must carefully take into account inherent differences between the two time-scales.

Correspondingly, the difference in the user base between the web and mobile devices may be huge, yet it is not easily predictable – another lesson learned. Being able to accommodate highly volatile user bases is a challenge to the system dynamics. In our case, for example, the system should not have discarded stories so eagerly (see Fig. 1), having so few storytellers. However, the design was based on the assumption that there are far more web users than players in the field, and thus the stories should have been in abundance. In the future we are determined to increase the system flexibility to adapt to situations like this.

The web site included a questionnaire which was answered by eight web players. Apparently the players who answered the questionnaire had generally enjoyed the experience, so it is difficult to make any conclusive remarks. To exemplify nature of the results, we include here answers for three of the questions:

Any surprise elements? “yes, when it became possible to be funny and make connections ... very surreal”, “Yes, when I saw the image that was for the words I selected.”, “discovering own sentences in other stories was very surprising.”, “It was a surprise that ”the game” didn’t seem to work at all. Also the term ”game” is misleading. It is more comparable to a mobileentertainment solution. Not into a game, as you don’t play it anyway.” **What did you like most?** “crafting the story. and wondering what photos would come out. using my knowledge of the city to make puns”, “Seeing everyone else’s creative efforts.”, “How people used my story und how the pics where done to my keywords.”. **Do you want to play this game again?** “Now that I have the hang of it, I’m thinking of new ways to play it...”, “yes! all the time”, “Yes. I would like to master whta I dont understand”, “Yes, please!!!!! It could lead to addiction, cause it is so much fun to write stories to the topics you offered and to wait how they will develope.”, “No, as it is not a game. If it’d work, I would reconsider writing in there.”.

5.3 Questionnaire

Qualitative data reflecting the user experience was evaluated with a questionnaire, which was collected immediately after the experiment. The questionnaire contained 26 statements on which the users were asked to answer on a 5-point scale between 1 (disagree completely) and 5 (agree completely). In addition there were 23 open-ended questions. A few individual users and 2-3 groups of users were also video interviewed after the experiment. Observation was carried out during the experiment with a video camera.

Of 184 players who participated the game, 99 returned a filled in questionnaire (56 males and 43 females). The age distribution was as follows: 24 players of age 18-24, 64 of age 25-34, and 15 of age 35+. Each player got an invitation to the Story Mashup evening party upon returning the questionnaire. In the following, we adopt the following notation for brevity: "Statement" (X), where X denotes the average of responses of 99 players on the 1-5 point scale.

As a general observation it was very clear that people really enjoyed playing the game: "It was fun and engaging to play this game" (4.07). This underlines that creating an interactive, engaging experience is key for a successful user participation for any large scale research experiment.

Of all the questions, it was a surprise that exactly the two major tasks that street players had to perform, got the highest rating: "I enjoyed hunting for photos" (4.63) and "I enjoyed the guessing part of the game" (4.39). This can be seen as a strong success from the game design point of view. It shows that players made an intellectual and emotional investment which Ryan [17] claims to be a precondition of an interactive medium to open its world to the user. Clearly, the use of mobile device as an interaction device and as an image capturing device in the context of our game is strong: "When I was holding the phone, I felt confident hunting for images and doing the guessing part" (4.18).

It seems that hunting for points and competing with other players in game context indeed provides a motivated activity. We can see from the data that the game mechanisms played a major role in getting players engaged and motivated. The ratings are very hig: "I like the competitive style of play" (4,32) and "I was motivated by the scoring mechanism" (4,25).

People were very active and also various interactions between people took place, which can be easily seen in the photos as well (see Section 5.1). People enjoyed socializing and teamplay: "I did prefer to play this game alone rather than joining a team of other players" (1.43). Players also felt, with some variation, that they are part of a joint authorship: In this case it was by contributing images to the web stories and ultimately to the large public display: "While playing I felt I was part of a joint activity between players on the web and street players in Manhattan" (3.22), "I felt I belonged to a joint, collaborative action contributing to a common goal" (3.59) and "For me it was an important part of this game to see the illustrated stories at the public display at Times Square (3.28)".

Also the usability goals were met, meaning that the system should be efficient to use, easy to learn and easy to remember: "The mobile application was easy to

use" (3.19), "It was easy to understand the game concept" (4.21) and "At any given moment it was clear to me what I was supposed to do" (3.51).

6 Conclusions

The game was a greater success than we expected. The game design proved to be engaging which supports our core design decisions. We were delighted by the amount of imaginative photos taken during the game. On the other hand the game was almost too motivating, since the data turned out to be quite homogenous, although we expected that some of the players would play rather lazily.

We think that the following factors contributed the most to the success of MSM: **Lazy Shooting**: A user with a camera is a powerful way to collect interesting data from the physical world. In addition, in our case it was effortless: The camera activated and sent the image automatically. This encouraged players to shoot many photos. **Ambiguity**: Players felt that clever and imaginative thinking contributed to their ranking. Players spent most of their time doing something that is natural to human beings: Trying to infer other people's intentions or trying to infer how other people would infer my intentions. **Teamplay**: Players formed teams spontaneously and used them to act out difficult words. Teamplay also lowered barrier for crossing the boundaries of public behavior. **Speed**: The game felt immersive since the players did not have time to think about anything else. Since the players' ranking was updated all the time, even a short pause resulted in a noticeable drop in score. **Integrated Game-Flow**: A single simple game client was used to play the game. Even the camera was included in the client. This ensured that we were able to streamline the UI to the bare minimum. With rapid prototyping tools like Python for S60, one does not have to accept suboptimal interfaces.

On the other hand, we could not find satisfying answers to the following issues: **Time and Web**: Even though technically our web site worked as expected, we would have expected more web players participating in the game. We assume that a major reason for the smallish number of visitors was that the game interface was open only during the game, for some 1.5 hours. In the future, we have to pay more attention to matching the physical and the virtual timescales. **Scalability**: Organizing a game involving almost 200 physical players was a major feat. The semi-manual approach for installing the phones does not scale easily to even larger settings. **Duration**: The game lasted for 1.5 hours which is enough for an intensive game but in order to get more data, we would like to keep the system running for a longer period of time. **Adaptability**: If the game involves web or it is otherwise open for anyone to participate, the rules should adapt to varying number of players and activity. This is especially important if the game runs for a longer period of time. Many of the previous issues relate to the natural friction between the web and the physical world. It is clear that further experiments are needed to find best practices and design frameworks for successful interaction between the two worlds.

References

1. Brown, B., Chalmers, M., Bell, M., MacColl, I., Hall, M., Rudman, P.: Sharing the square: collaborative leisure in the city streets. In: Proc. of the Ninth European conf. on Computer-Supported Cooperative Work, Springer (2005) 427–447
2. Benford, S., Magerkurth, C., Ljungstrand, P.: Bridging the physical and digital in pervasive gaming. Communications of the ACM **48**(3) (2005)
3. Bamford, W., Coulton, P., Edwards, R.: A massively multi-authored mobile surrealist book. In: Proc. of the SIGCHI conf. on Advances in computer entertainment technology, ACM Press (2006)
4. Brignull, H., Rogers, Y.: Enticing people to interact with large public displays in public spaces. In: Proc. of INTERACT-03. (2003) 17–24
5. Scheible, J., Ojala, T.: Mobilenin combining a multi-track music video, personal mobile phones and a public display into multi-user interactive entertainment. In: Proc. of the ACM conf. on Multimedia, ACM Press (2005) 199–208
6. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2004) 319–326
7. Benford, S., Crabtree, A., Reeves, S., Sheridan, J., Dix, A., Flintham, M., Drozd, A.: Designing for the opportunities and risks of staging digital experiences in public settings. In: Proc. of the SIGCHI conf. on Human Factors in computing systems, ACM Press (2006) 427–436
8. Falk, J., Ljungstrand, P., Bjork, S., Hannson, R.: Pirates: Proximity-triggered interaction in a multi-player game. In: Extended Abstracts of Computer-Human Interaction (CHI). (2001) 119–120
9. Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., Brown, B., Rowland, D., Benford, S.: Interweaving mobile games with everyday life. In: Proc. of the SIGCHI conf. on Human Factors in computing systems, ACM Press (2006) 417–426
10. Chalmers, M., Bell, M., Brown, B., Hall, M., Sherwood, S., Tennent, P.: Gaming on the edge: using seams in ubicomp games. In: Proc. of the SIGCHI conf. on Advances in computer entertainment technology, ACM Press (2005) 306–309
11. Thoresson, J.: Photophone entertainment. In: CHI'03: Extended abstracts on Human factors in computing systems, ACM Press (2003) 896–897
12. Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., Row-Farr, J.: Can you see me now? ACM Trans. Comput.-Hum. Interact. **13**(1) (2006) 100–133
13. Reeves, S., Benford, S., O'Malley, C., Fraser, M.: Designing the spectator experience. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2005) 741–750
14. Aoki, P.M., Woodruff, A.: Making space for stories: ambiguity in the design of personal communication systems. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2005) 181–190
15. Fellbaum, C., ed. In: WordNet: An Electronic Lexical Database. MIT Press (1998)
16. Laurila, J., Tuulos, V., MacLaverty, R.: Scripting environment for pervasive application exploration on mobile phones. In: Adjunct Proc. of Pervasive 2006. (2006)
17. Ryan, M.: Narrative as Virtual Reality. Johns Hopkins University Press (2001)