3.4.2 Multi-core optimization

The placement of the words is performed in a placer thread which runs separately from the rendering thread. The placer thread takes the list of the words and places them while ensuring that none of them intersects with the words already placed on the canvas. Using a separate thread for words placement is to make sure that the heavy computations in the placement logic do not cause delay in the rendering loop where the animation is handled. The placer thread is initialized once and is never destroyed until the program is terminated because the overhead of destroying and recreating the thread every time necessary could be costly. Instead, it waits indefinitely until it detects words that need to find new locations. The main reason for this is that, in ManiWordle, we may invoke the placer multiple times a second during the runtime to allow real-time manipulation of a word’s location, angle, and font, in which we need to push the words that intersect the updated word. The placer thread also spawns multiple collision test threads depending on the number of cores or logical threads available on CPU. The number of threads may vary depending on the system.

For the sake of the experiment, we used six threads. We ran the simulation multiple times and picked the number which showed the fastest computation time on the system used for the experiment, which is described in Section 4. The example of how the placer thread operates is shown in Fig. 4. Since multiple points are being checked for availability simultaneously, we can assign one thread per processor core and shorten the time approximately by the factor of the number of threads.

3.4.3 Collision detection with reduced-resolution image

It can be quite expensive to do the collision detection on spline-based shapes (one of vector-based collision detections in which the modern fonts are defined). Instead, ManiWordle makes a 1-bit mask image (literally, a bitmap) for each word. For every pixel on which the image of a word has a non-transparent value, the mask image will have a bit 1, and 0 otherwise. The reason for using a 1-bit mask image is to maximize cache hits. One pixel in an image of a word is normally represented in ARGB format which is 32 bits. However, the actual color of the word is irrelevant when checking for the collision. Therefore, we can reduce the data into a 1-bit-per-pixel image, resulting in much better cache hits when loaded into memory.

ManiWordle reduces the size of the mask image to have only one third the original size in each dimension, resultanty having only one ninth of the number of pixels compared with the original mask image. Collision detection based on these smaller mask images does not perform as accurately as either of vector-based collision detection or pixel perfect test (with full resolution) that guarantees no overlaps between any two given words; however, it is much faster due to the reduction in the number of pixels to check to the only one-ninth the number of pixels in the original mask image and it is shown to produce the results that are good enough for the purpose of ManiWordle with minor overlap in a few places.

By exploiting these optimizations, specifically the usage of multiple threads, image representation by 1-bit mask, and reduction of the mask image resolution to 1/9, we can get a significant speed boost. When these optimization techniques and the aforementioned heuristics (i.e., the adjusted growth rate of spiral radius and the intersection test with some interval) are all combined, ManiWordle delivers the animated rendering of 500 words with 30 frames per second. It is fast enough to allow users a real-time interaction with ManiWordle running on desktop computers including the one used in the experiment, which is described in Section 4.2.5.

4 Evaluation

To evaluate ManiWordle, we conducted two studies. We first conducted a preliminary usability study to identify major usability issues and investigate users’ reaction to ManiWordle. After making some improvements according to the results from the first study, we ran a controlled experiment to see if ManiWordle with more flexible user control is better than Wordle in terms of subjective satisfaction for creating word clouds.

4.1 Usability Study

We recruited 6 participants (3 females and 3 males) who were all graduate students majoring in Computer Science. None of the participants has used Wordle before. After 15 minutes of training, participants were asked to manipulate position, orientation, and typographical properties of several big words to match a target layout presented. The big words that participants had to manipulate were highlighted with an arrow mark in the target layout image printed on a paper. (The target layouts were prepared using ManiWordle.) They repeated the task three times with three different data: 1) a Wikipedia entry on Yu-Na Kim, a gold medal-winning South Korean figure skater of the 2010 Winter Olympics; 2) a Wikipedia entry on StarCraft, a popular strategy video game; and 3) the Wordle paper by Viegas et al [23].

All participants finished the task with each input text within 5 minutes without any difficulties. Participants seemed enjoying using ManiWordle. Two of them commented that “it was fun,” and three of them said that “it was intuitive.” We did not identify any major usability issues from the study.

4.2 Controlled Experiment

4.2.1 Datasets and the Task

For the study, three different text datasets were prepared that varied regarding the emotional attachment between the text and the participants. This is based on the finding that 57 percent of the Wordle users are the actual authors of the text, while only 7 percent has never looked at it before [23]. The first text with the least emotional attachment was an InfoVis conference paper about Wordle written by Viegas et al [23]. None of the participants has read the paper before. The second text with moderate emotional attachment was a Wikipedia entry on Yu-Na Kim. This article was selected considering her great popularity and good reputation among South Koreans. The last text with the most emotional attachment was the participants' own paper. We removed the “references” section from each paper to prevent too many non-keywords competing for the word counts. We also prepared a text for the practice task. A Wikipedia page on Beatles was used because it was familiar to the participants. Each participant was asked to make as aesthetically appealing presentation as possible for each of the three text datasets.

4.2.2 Participants

We recruited 12 participants (11 males and 1 female). 11 of them are majoring in Computer Science and one participant is majoring in Chemical Engineering. They are all graduate students. None of the participants has used Wordle before, while three of them have seen other tag clouding techniques on internet blogs. We screened participants so that they have an experience in writing a conference or journal paper in English. Participants were asked to email the experimenter or bring the favorite of their own papers written in English in their USB thumb-drives. They were given about $15 for their participation.
4.2.3 Hypotheses
We hypothesized that the extended functionality of Wordle results in higher user satisfaction and it lets people feel more creative while making their presentations. We also expected that the more the participants are emotionally attached to the text, the more effort they are going put in creating the word cloud.

4.2.4 Study Design and Procedures
We ran the study as a 2 (Visualization: Wordle, ManiWordle) x 3 (Text: the least, moderate, and the most attachment) within-subject design. Each participant performed the task (i.e., making his/her aesthetically appealing presentation) for all the three text datasets using all the visualizations. We counterbalanced the order of visualizations to avoid the learning effect. Within the six participants testing the two visualizations in the same order, the three text datasets were also counterbalanced.

Before the experiment, participants were given a tutorial on the experiment. Before beginning the test on each visualization, participants also performed a representative practice task in order to familiarize themselves both with the task and visualization. For each task, participants were told that they have roughly 5 minutes to complete. After 5 minutes, they were told to try to wrap up in 1 minute. However, they were allowed to spend more time as needed if they were still not satisfied with the result. The 5 minute limit was given to prevent participants from spending too much time on the first few tasks and then being exhausted for the rest of the experiment. After each session with a visualization using all the three text datasets, participants were asked to fill out the post-session questionnaires for subjective evaluation. The same procedure was repeated for the other visualization. When a participant finished both sessions, they were asked to fill out a demographic survey along with a question which asks their preference between the two visualization tools and the subjective reasons. The experiment took about an hour.

If a participant was exposed to Wordle in the first session, we only explained the features added to ManiWordle before beginning the second session using ManiWordle. For the opposite case where participants tried ManiWordle first, we were very careful not to use negative terms like “restriction,” “removal,” “restrain,” or “limit” when explaining Wordle in the second session. We did not want participants to feel they are given the additional functionalities of ManiWordle first and got them taken away for Wordle.

4.2.5 Experiment Setup
Each participant worked on a quad-core PC with a 27" LCD widescreen display running at a 1920x1200 pixel resolution. The system was also equipped with a NVIDIA 9800 GX2 GPU with 512 MB of memory. The program was maximized to fit the entire screen except for the taskbar. All events were logged by recording every single user interaction such as re-color, re-layout, change color/font, etc. For statistical analysis on the number of user interactions, we counted every user interaction and the interactions that affected only the global layout such as randomize, re-layout, remove a word, etc. When participants had made a satisfied aesthetic presentation using visualization, they were asked to notify the experimenter the task completion. The task completion time was manually measured.

4.2.6 Results
We performed statistical analysis on participants’ responses to the six questions from each session, preference data from the post-study questionnaire, and elapsed time to finish each task. After each session, participants were asked to answer the following 6 questions to collect subjective ratings about each visualization by using a 7 point Likert scale [Rating: 1 = Strongly disagree; 7 = Strongly agree] (Table 1). The questions with significantly different ratings are marked with an asterisk (*).

We analyzed the subjective ratings using Friedman’s Chi-Square test. Participants were significantly more satisfied with the result layout of ManiWordle than that of Wordle ($\chi^2(1) = 9, p = .039$).

Other than that, we did not find any statistically significant difference between the visualizations.

In the post-study survey, we did not find any statistical significance between the order of the visualization participants tried and the preference. 10 out of 12 participants said that they liked ManiWordle more than Wordle. Every participant who preferred ManiWordle mentioned that its ability to fine-tune the layout and the flexibility of manipulation was the reason. One participant specifically stated that it was easy to fix big words in the location he desired and fill out the rest through the automated process. Another participant said that while ManiWordle was more difficult to learn and some of the rules on conflict resolution were confusing, it was still better than Wordle which does not allow the fine tuning. Also, two participants mentioned that the animation makes ManiWordle less boring and more fun.

The two participants who preferred Wordle to ManiWordle mentioned that they did not like to adjust too many things. One of them criticized that the results of two programs were not quite different. And he also said, “I had some feelings that if I changed something, I couldn't get it back later.” He did not utilize “undo” which could have eased the problem. When asked why, he said he forgot that it was there. The other participant said that ManiWordle requires too much effort and labor. He said that Wordle requires much less time to build a word cloud and often the result was as good as that of ManiWordle.

We also investigated the effect of the visualization and the emotional attachment level on the amount of time to complete a task. For each task, we measured the total elapsed time to make a satisfying tag cloud with each visualization. We ran a 2 (Visualization: Wordle, ManiWordle) x 3 (Text: the least, moderate, and the most attachment) analysis of variance and Tukey’s HSD post-hoc test. We found a significant main effect of Text ($F_{2,66} = 3.42, p = .039$) with post-hoc tests showing that participants spent more time to create a wordle for their own research paper than for the Wordle paper ($p = .03$) (Fig. 5).

We recorded all of the user interaction to analyze the effect of the visualization and the emotional attachment level on the number of user interactions in creating Worldles. We performed a 2 (Visualization: Wordle, ManiWordle) x 3 (Text) ANOVA test with the number of total interactions as the dependent variable. We found a significant main effect of Visualization ($F_{1,66} = 5.01, p = .03$). This result shows that participants initiated significantly more user interactions with ManiWordle than Wordle. We performed another 2 (Visualization) x 3 (Text) ANOVA test with the number of only the interactions that affected the global layout as the dependent variable. Again, we found

![Fig. 5. Task completion time (in average) for three text datasets. Error bars represent standard error.](image344x78 to 537x187)

Table 1. Subjective Responses to Six Questions (Average Ratings)

<table>
<thead>
<tr>
<th>Questions</th>
<th>Wordle</th>
<th>ManiWordle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: It was easy to learn this visualization.</td>
<td>5.54</td>
<td>5.23</td>
</tr>
<tr>
<td>Q2: It was easy to use this visualization.</td>
<td>5.31</td>
<td>5.38</td>
</tr>
<tr>
<td>Q3: I liked to use this visualization.</td>
<td>5.38</td>
<td>5.62</td>
</tr>
<tr>
<td>Q4: It was fun to use this visualization.</td>
<td>5.46</td>
<td>5.77</td>
</tr>
<tr>
<td>Q5: I felt creative while using this visualization.</td>
<td>5.08</td>
<td>5.54</td>
</tr>
<tr>
<td>Q6: Overall, I am satisfied with the result layout. *</td>
<td>5.31</td>
<td>5.77</td>
</tr>
</tbody>
</table>

Fig. 5. Task completion time (in average) for three text datasets. Error bars represent standard error.
a significant main effect of Visualization ($F_{1,66} = 5.40, p = .02$). This result shows that the participants performed significantly less user interactions to globally change the layout with ManiWordle than with Wordle.

4.2.7 Observations on the final layouts

Using ManiWordle, most participants produced the layouts which could have not been produced using Wordle. For example, one participant emphasized some words using a color-change feature in ManiWordle (Fig. 6 left) Words for Yu-Na Kim’s name, job, and victory in the world championship are significant keywords which can only be understood if the participant knows theme and context of the text. Finding out these true keywords instead of just using the number of appearance is a much more challenging problem for the automated algorithms. Fig. 7 left shows results from a participant who clustered the words based on the semantic meanings in ManiWordle. He performed clustering for all three input texts. This type of clustering is also a challenging problem and requires users’ involvement. ManiWordle utilizes users’ knowledge in the word cloud generation process, therefore produces the layout easily appreciated by users.

5 Discussion and Future Work

The results of the controlled experiment support the need of flexible user control provided by ManiWordle. First of all, ManiWordle’s extended control over Wordle yielded higher user satisfaction. Also, most participants preferred ManiWordle overall compared to Wordle. In fact, they significantly utilized the added control for the individual words. Among all the configuration changes participants made with ManiWordle, 46% of them were to tweak the individual word configuration on average. Furthermore, participants changed the global configuration significantly less frequently with ManiWordle than with Wordle.

It is well known that usability of interactive systems decreases as the system becomes more flexible by supporting more functionality [10]. While ManiWordle’s interaction was relatively simple, it is still adding some complexity to Wordle. Nonetheless, we found it very encouraging that ManiWordle was as easy to learn and use as Wordle based on Q1 and Q2 in Table 1, where we did not find any significant difference.

The study results also supported our hypothesis that participants would spend more effort in creating a word cloud for the text that they are more emotionally attached to. There was no significant effect of visualizations on time spent to create the output, indicating that people still spend a considerable amount of time with Wordle even if they did not have control. Interestingly enough, we did not find any effect of text on the number of user interactions. This may imply that participants had more cautious interaction for the text they cared about more.

Our work was originally, in part, inspired by the fact that people felt creative in using Wordle. Given that people felt creative without much control, we initially anticipated that the flexibility provided by ManiWordle may result in users’ feeling more creative. To our surprise, participants neither thought that it was more fun to use nor felt more creative even if they had more control with ManiWordle. This leads us now to wonder if people might feel creative as long as they have aesthetic results since they are inclined to have more positive attitude toward aesthetic designs than less-aesthetic designs [9]. We also suspect that it resulted from the fact that all participants were computer science graduate students. Thus it might be interesting to see if other user population with artistic aptitude (e.g., graphics designers) would take flexible ManiWordle as a more creative way to design a layout.

On the other hand, the encouraging results of the controlled experiment opened several avenues for future research. In this study, since we wanted to preserve the original design of tag clouds, we did not allow people to change the font size. Given that a significant number of people created wordles not specifically for data analysis and many people manipulate the text to be fed into Wordle to change the font size, we wonder if it is better to allow people to fully manipulate the Wordle output. In addition to letting people directly manipulate the font size, we can allow them to add words to the
word cloud. It would be interesting to investigate whether enabling richer user control could change people’s reaction in terms of fun and creativity in future work.

Also, as some participants have commented, ManiWordle can be extended to support simultaneous selection of multiple words. This will allow users to manipulate the contextually relative words together. For example, users may set the words with similar colors, align them on common horizontal, or form a new cluster in somewhere distinctive (Fig. 7).

In terms of the layout algorithm, it would be interesting to employ different physics-based layouts to simulate the words as blocks on the table. We can also try to incorporate other context-aware optimizations, such as an automatic approach to context-preserving dynamic word clouds introduced by Cui et al. [26].

Another promising avenue is exploring more natural user interfaces with which people can interact with ManiWordle. For example, manipulating words with fingers/gestures using a multi-touch screen may also increase people’s satisfaction in terms of fun and creativity.

6 Conclusion

Wordle creates aesthetic visual representations, fusing together with text/document visualization techniques, and has attracted a huge number of people to participate in creative wordling feasts. In this paper, we have presented a Wordle-based visualization called ManiWordle, which enables custom manipulations to revamp interactions with the layout. ManiWordle provides flexible control over Wordle by allowing people to directly manipulate typography, color, position, and orientation for the individual words. We have described our design rationale along with the interaction techniques for tweaking the layout.

We conducted a preliminary usability study to identify major usability issues and areas of improvement. Through the controlled experiment, we compared ManiWordle to Wordle in terms of user satisfaction for the layout result as well as how easily people could learn and use. The results suggest that ManiWordle’s ability to provide flexible control over Wordle yields significantly higher user satisfaction without introducing more difficulties. Also, most participants liked ManiWordle more than Wordle.

Acknowledgments

We would like to thank HCIL & VPRG members for helping us with valuable comments and suggestions during design and evaluation of ManiWordle. This work was supported in part by the BK21 Project in 2010 and by Basic Science Research Program through NRF of Korea funded by the MEST (No. 2010-0002266 and No. 2010-0017178). The ICT at Seoul National University provides research facilities for this study.

References