Exploring Input Modalities for Collaborative Storytelling

Zach Havens (7671770), Raquel Thiessen (7813231)
Department of Computer Science
University of Manitoba

ABSTRACT

Joint storytelling is a collaborative activity with multiple benefits, such as improved language development for children. Many studies have created AI-supported tools for collaborative storytelling with varying input mechanisms. Very little research has compared the differing quality of different input methods. We built a collaborative storytelling prototype with multiple input methods to compare which method works best and engages the children.

KEY WORDS

Human-Computer Interaction, Child-AI Interaction, Storytelling-AI

I INTRODUCTION

Storytelling is a frequent activity that children engage with in their childhood. Storytelling has multiple developmental benefits, including improving a child's communication and linguistic skills [1]. Complex storytelling can be challenging for a parent to facilitate, but we can provide children with further opportunities for creative and beneficial storytelling using current technology.

Previous interactive storytelling prototypes have shown to be beneficial to a child's development. In StoryCoder [2], they had children listen and then modify the stories. Their results showed their system effectively helped develop the children's computation thinking skills. Robot prototypes have been developed to facilitate creative play through storytelling [3]. Similarly, research has looked into using conversational agents such as Amazon Alexa to support a child's literacy development [4], language acquisition [5] and science learning [6].

Many of these prototypes use speech as the input. Some due to the technical limitations, and others simply use it without much explanation. We want to investigate what input methods would best engage the child in the storytelling process. Specifically, we are looking at the following research questions:

R1.1) What interaction method best supports child-AI collaborative storytelling?

- **How might age and other developmental factors impact this on an individual level?**

II RELATED WORK

Joint storytelling is a collaborative activity with multiple benefits, such as improved language development for children. To justify using AI technology for facilitating creative play, we must understand the media equation. Reeves and Nass [7] introduced the media equation stating that media equals real life. They argued that we subconsciously treat it realistically no matter how much we know or think about it. Multiple studies support this for characteristics such as motion, emotion, social roles, and more. This research provides evidence for the argument that people treat technology such as AI similarly to real life, no matter how unreasonable. Knowing this, we can justify that an AI intervention could be an effective way to

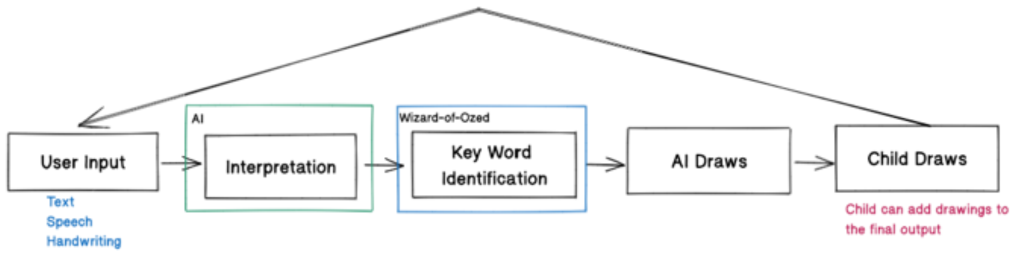

Figure 1. The Collaboratory Storytelling System

facilitate creative play. We do not consider it a replacement for playing with human peers.

Digital tools for facilitating collaborative storytelling have been investigated for a long time. A notable early study in this space was KidPad, an interface that allowed multiple children to tell and illustrate a story together [8]. Given the lack of AI technologies at the time, the focus was primarily on interface elements and drawing tools to improve experience and engagement. Since then, AI has been integrated into storytelling in many ways, such as generating questions to help increase comprehension [9], suggesting narrative elements for human collaborators [10], or allowing a robot to listen to a story told and react with gestures and other motion [11].

The primary inspiration for this specific project was a study and system called StoryDrawer [12]. It is a collaborative storytelling system with two main approaches. The first is a child says, and an AI draws a model where the child tells stories verbally, and the web application transforms the oral description into drawings that the child can elaborate on. The second is a child scribbles, and an AI completes. In this version, the child draws an illustration that is analyzed by a drawing recognition model to create a textual representation. This representation is passed to a narrative model to further the story, and drawings for the new story elements are rendered on the same canvas as the child's original contributions. Overall, this study illustrates the value of both voice and drawing as inputs for a collaborative storytelling system but does not compare the two, instead of focusing on overall engagement and enjoyment. Another contribution made by StoryDrawer is the inclusion of a button that presents a set of story ideas to the child to minimize problems getting started, which users seemed to value, indicating that they would like such a feature in other systems.

Our literature review highlighted there are few, if any, current studies that actively compare different input methods for child-AI interaction. Prior research has investigated specific input methods and how to improve them, such as speech recognition [13], text-based inputs [14], and drawn images [12]. Each of these input methods has been independently shown to be effective, but they have not been compared to one another in a collaborative storytelling setting.

It is worth noting the research that outlines many potential ethical issues when using AI with children. These issues include surveillance, privacy, security, and inclusivity

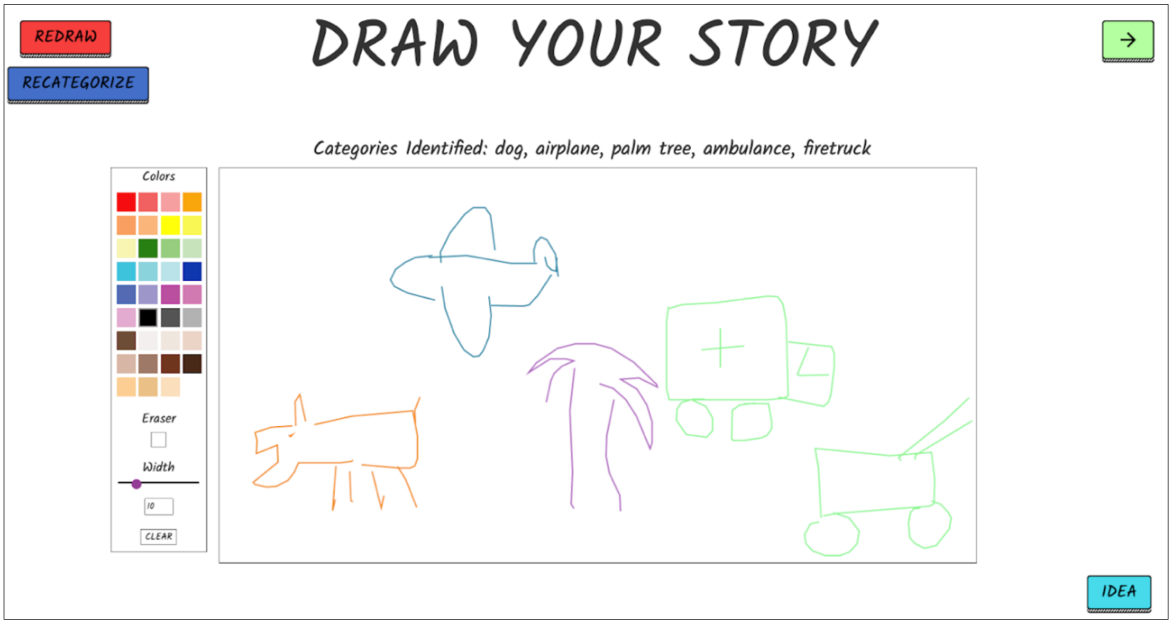

Figure 4. The Canvas Interface

[15]. As a result, we must have a transparent design, education, and regulation of these agents when used with the public. Research suggests that children do not demonstrate awareness that the technologies are recording their interactions. Parents have expressed specific concerns about recording and monitoring their children's activities and what data is held by companies. Though these ethical issues will not be directly applicable to our prototype, they were front of mind as we designed our study.

III DESIGN PROCESS

Our design processes consisted of the following five steps:

1) Literature Review
2) Paper Prototypes
3) Refine Prototypes
4) Interaction Storyboard
5) Build Web Prototype

We started our project with a general literature review to understand what similar projects have been done, what their findings were and what they believed could be approved upon. Once we decided the focus of our project, input modalities, we made some paper prototypes of what we thought our collaborative story drawing system could look like. Following this, we chose the prototypes that we liked best and refined them as shown in *Figure 2* and cleaned them up using basic digital tools.

We also created basic interaction story boards and diagrams as shown in *Figure 3* so we would understand the entire process and the different steps for the different versions. Lastly, we used our story boards and digital prototypes to build out web prototype.

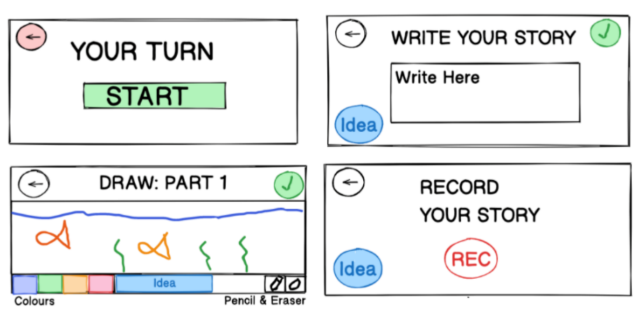

Figure 2. Refined Prototype Examples

IV PROTOTYPE

IV.I Input Types

The first input type we used was plain text. For this input type the child merely had to type out their story and then submit it. This was intended to act as a baseline without any interpretation phase. However, during our pilot the participants used an iPad to interact with the prototype which allowed them to use the on-screen keyboard for input. This keyboard option has completion and correction suggestion based on partial inputs from the user, meaning that there was a light-weight interpretation mechanism available to the user although its use was optional. The second input type was handwriting. In this input type, the child had to write their story on a digital notepad. The MyScript library [16] was employed to interpret the handwriting into written text. Lastly, we had the speech input type. The child would record their story orally. We presented the child with the option to replay and rerecord their input before submitting it. We used the Web Speech API built into modern browsers to interpret the speech into text as well as synthesize speech for validation. Drawing as input (as in StoryDrawer) was also considered but was omitted due to a lack of access to pre-trained models, and limitations on our ability to train our own.

IV.II Canvas

The most important component of our prototype was our canvas (*Figure 4*). Using the generated interpretations of the story, our *Wizard-of-oz* would select image categories that the AI should draw as the starting outline for the drawing of the story. With these categories our system will gather images for these categories using the "Quick, Draw!" dataset [17] and randomly position them on the canvas. Following the generation, the child is given an opportunity to add what they want to the image representing their story.

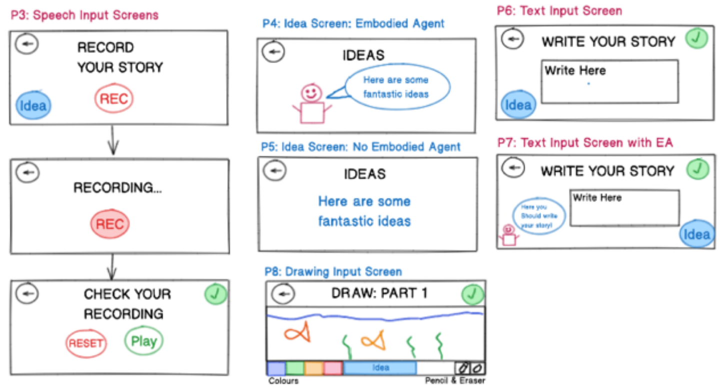

Figure 3. Storyboard Examples

IV.III "Quick, Draw!" Dataset

We used the "Quick, Draw!" [17] dataset for our project. This dataset contains thousands of open-sourced human-drawn doodles in each of 345 categories. These images were collected in a gamified training data collection mechanism for Google's image recognition neural network. In our implementation we stored 100 random images from each category on the webserver. A web endpoint was provided that returned a random doodle from the specified category. These doodles were used to compose the illustration for the child's story. One advantage of these drawings is that they are all quick sketches instead of well-drawn artworks, providing lots of opportunity for elaboration on top of them. Quick sketches can help establish that these illustrations can be fun and simple. There is a limitation on the number of categories available. We felt it to be an appropriate trade-off when compared to having a hand-draw

Pseudonym	Version(s)	Engagement	Preference / Comments	Time Writing	Time Drawing
Peach	All versions	Relatively low engagement throughout	Expressed they did not like the speech but did not have a preference otherwise	1 min (avg)	1.5 min (avg)
Toad	Text	High Engagement in story writing, lower in drawing	Typing – they said they wanted to type before we even started	14 min (total)	25 min (total)
Luigi	Speech and Handwriting	Fairly decent engagement, had simple stories	Struggled with speech so we ended up switching to handwriting	21 min (total)	21 min (total)
Bowser	Handwriting	High Engagement throughout	Started on handwriting and they went slowly through the drawing	25 min (total)	16 min (total)

Table 1. Participant Information

images during the study. StoryDrawer [9] used the same dataset, and they dealt with similar trade-offs.

IV.IV Idea, Redraw and Recategorize

We had three primary buttons on our canvas page that the user could use. The first inspired by the StoryDrawer prototype, was the "Idea" button. This button was to help children if they were stuck or needed inspiration for their story or drawing. The second button was the redraw button. If the child did not like the canvas generated, they could ask the AI to try drawing again. Another alternative was to clear the canvas and purely draw the image themselves. They could press redraw as many times as they wanted. Lastly, we had the recategorize button. We told the children that this button would ask the AI to think of better categories for their story. In reality, it really just asked the wizard to choose new categories.

V STUDY METHODS

V.I Participants

We ran four pilot studies with children in first grade. For each pilot study, there were three stories written by the participant. For each study, the participant used a different input method. Each input method had four rounds of input and drawings. We changed the order of the input methods to balance any learning effect.

V.II Study Session

Our study started with a brief introduction to the process and a demo of the system. We originally planned to have each child go through all three sessions trying the different input methods, however the children took a lot longer to do the sessions than initially anticipated. As a result, we had children just using one system and did not have an interview but instead used what they used throughout to try to understand their general opinion of our system.

V.III Data Collection

During our study we collected multiple pieces of data to try to best understand the process and engagement of the child during the collaborative storytelling. We collection data of the original input and the AI interpretation to look at the accuracy of the AI for interpreting the child's input. We saved the originally generated

AI image and the final image to see how much the child added or changed. Lastly, we collected button press and time data to help understand how long the child took in different components and whether they utilized the redraw, recategorize or idea buttons.

VI RESULTS

VI.I Input Engagement

We found that Toad had the longest and most in-depth story. However, the child expressed a lot of interest in story writing and so the results might be skewed. All the children seemed to like the handwriting over the speech, however the handwriting had limited space for writing a story. Both children who did speech, struggled to use it. This may be partially because the process was not super intuitive, but also, they could not easily change parts of the story and instead had to re-record. The children also thought of their stories slowly and the system stops recording when it hears a long pause. Of all the methods, the handwriting input seemed to lead to the most language learning. The text input had auto complete, and speech just required them to talk. In the handwriting version they were writing words and we worked with them on sounding out words to help them spell words for their stories.

VI.II Canvas Engagement

The canvas engagement varied quite a bit between the children. Bowser, for example, spent a majority of the time on the canvas and added to the images generated. Luigi spent an even amount of time drawing and writing their story. When Luigi drew the picture, they often cleared the canvas and drew their own. Toad also cleared the canvas as drew their own picture, but theirs did not seem as related to their story and contained a lot of words. Peach was not interested in drawing and instead often pressed the redraw button repeatedly until they liked the image and then moved onto the next round.

Figure 5. Bowser's Images

VI.III AI Accuracy

Unfortunately, this is an aspect of the prototype's performance that we were unable to effectively exercise due to a bug in the prototype. The intent was for the system to perform input interpretation client-side and log the interpretation such that it was presented to the wizard for categorization. During the study, a previously unexposed bug caused many of the interpretation logging operations to fail, and many of the categorizations were done based on the wizard's direct exposure to the child's inputs (as opposed to from the AI interpretations). However, the limited evidence that we were able to collect showed that speech and handwriting were capable of equivalent accuracy, but frequently failed to correctly interpret the inputs. Their success depended heavily on the clarity of inputs, which for children of this age had varying levels of development across both input types. In our prototype, handwriting seemed to hold an advantage in that the handwriting panel allowed for revision and modification where a child could iteratively improve the legibility of their handwriting as they continued to write, whereas the speech input required the child to re-tell the entirety of their story every time they were unsatisfied with the interpretation, allowing for more deliberate inputs.

One surprising result was the adoption and quality of the iPad/iOS's predictive on-screen keyboard. This was not initially considered as an AI-interpreted input mechanism, but in hindsight was likely the most accurate and helpful. The keyboard allowed the children to interact directly with the interpretation of their inputs, and often resulted in more expressive and error-free textual stories that we believe would allow them to be more recognizable by NLP/narrative models as well as help improve the children's spelling over time.

VI.IV Button Usage

The redraw button was by far the most used button. We logged over 300 button presses for the redraw. The children really liked seeing the different drawings the AI would come up with and would press redraw until they saw one, they liked. We logged 25 uses of the idea button. One child did use it to help inspire them, others just tested what the button did. We only logged 12 uses of the recategorize button, but some of these were used to send them back to a waiting screen when bugs appeared.

VII DISCUSSION

VII.I Interpretation

As a result of technical difficulties, we could not thoroughly investigate which input from a technical side would be more accurate or best to use. We did get some feedback. Peach (who used all the systems) said, "I don't think the voice does good." Similarly, Luigi started on speech but switched to handwriting because they struggled to use the speech system. From our pilot, we believe that speech is the simplest input method as speech generally comes first before typing and handwriting skills. However, it may not be the best for the overall engagement of storytelling. The handwriting and typing input methods that allowed for easy changes and review of the story seemed to engage the most in the story writing component and resulted in more accurate interpretations.

We were trying to test the use of the different inputs to find a potential "best" one, but the children seemed to perceive it more as the flexibility of the system. For example, Luigi said, "everyone gets to do what they like." After running our pilots, I believe giving all three options might allow us to adjust a system based on their age and individual characteristics. Having flexibility could let them focus on different goals for their creative play. We've shown that an implementation supporting multiple input methods is feasible with limited time investment given current AI libraries and frameworks, so we believe this to be a tractable design consideration.

VII.II Challenges

We had technical challenges while running our study and we had unexpected challenges trying to facilitate the storytelling process with the children. Some children did not care for the initial writing story portion and would write a simple sentence and then do elaborate drawings. Others spent a lot of time writing an elaborate story and did just a simple drawing. Our experience through this process made it clear that we must have a more flexible system that allows them to do storytelling in the way that fits best for them.

VII.III Limitations

One limitation of our project was our limited participant pool. As a result of the time frame of the course, we had a short time to run the study sessions. Ideally, we would have had multiple participants for different age groups. A better sample would have allowed us to get more evidence and see the potential influence of the child's age on the input type that worked best, and on which was most engaging. Another limitation was that we did have to have a wizard in our implementation due to technical constraints. We would have liked to have it fully implemented to test the realistic feasibility of this type of system. Our last limitation regards our drawing data set. Though large, a story can contain no categories. In that case, our AI would not be able to contribute at all to the canvas.

VII.IV Future Work

For our future work, we would address this limitation by adding a component allowing a child to add their images with an associated label that could be used by the AI when generating the canvas. Another feature we would want to add is including an embodied version of our AI. Our AI was more behind the scenes making it less obvious to the user that the collaboration was with an AI. Using an embodied agent would more clearly indicate that this contribution was by an AI. Having an embodied agent could potentially help with engaging the child. We also considered implementing drawing recognition as an input and then having the AI contribute to the drawing. For young children who cannot type, write, or potentially for those who are non-verbal, drawing recognition would allow them to still do storytelling collaboratively.

VIII CONCLUSION

We've developed a collaborative child/AI storytelling prototype that supports multiple types of input (speech, handwriting, and text) for the purpose of comparing the quality and accuracy of the differing input mechanisms. We presented the prototype to a group of first grade students with support from a *Wizard-of-oz* to mitigate the need for narrative analysis models while still providing a fun and complete experience for the children. Despite technical difficulties encountered with the prototype, we found that speech was the most natural input mechanism in line with its earlier development as a communication method in children. Interestingly, text and handwriting methods seemed to be more accurate as the children were able to slowly iterate and refine their input more easily with help foster understanding by the AI models. Each mechanism showed signs of encouraging improved communication in the children, and as such we suggest that collaborative storytelling system designs support multiple input mechanisms to encourage engagement and facilitate both play and learning for the broadest range of development and skillsets in children.

IV REFERENCES

[1] J. Peck, "Using storytelling to promote language and literacy development,"

Read. Teach., vol. 43, no. 2, pp. 138–141, 1989.

[2] G. Dietz *et al.*, "StoryCoder: Teaching Computational Thinking Concepts Through Storytelling in a Voice-Guided App for Children," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–15.

[3] L. J. Hubbard, Y. Chen, E. Colunga, P. Kim, and T. Yeh, "Child-Robot Interaction to Integrate Reflective Storytelling into Creative Play," *ACM Int. Conf. Proceeding Ser.*, 2021, doi: 10.1145/3450741.3465254.

[4] Y. Xu, S. Branham, X. Deng, P. Collins, and M. Warschauer, "Are Current Voice Interfaces Designed to Support Children's Language Development?," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–12.

[5] N. Bhatti, T. L. Stelter, S. McCrickard, and A. Kelliher, "Conversational User Interfaces As Assistive Interlocutors For Young Children's Bilingual Language Acquisition," in *ACM International Conference on Interactive Media Experiences*, 2021, pp. 208–211.

[6] Y. Xu, "Using conversational agents to foster young children's science learning from screen media," in *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*, 2020, pp. 14–19.

[7] C. Reeves, Byron; Nass, *The media equation: How people treat computers, television, and new media like real people*. 1996.

[8] A. Druin, J. Stewart, D. Proft, B. Bederson, and J. Hollan, "KidPad: A design collaboration between children, technologists, and educators," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 463–470, 1997.

[9] Z. Zhang *et al.*, "StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement," Feb. 2022, doi: 10.1145/3491102.3517479.

[10] M. Kreminski, M. Dickinson, M. Mateas, and N. Wardrip-Fruin, "Why Are We like This?: The AI Architecture of a Co-Creative Storytelling Game," in *ACM International Conference Proceeding Series*, 2020, doi: 10.1145/3402942.3402953.

[11] S. Al Moubayed *et al.*, "Multimodal Feedback from Robots and Agents in a Storytelling," 2021.

[12] C. Zhang *et al.*, "StoryDrawer: A Co-Creative Agent Supporting Children's Storytelling through Collaborative Drawing," in *Conference on Human Factors in Computing Systems - Proceedings*, 2021, pp. 1–6, doi: 10.1145/3411763.3451785.

[13] M. Kumar, S. H. Kim, C. Lord, T. D. Lyon, and S. Narayanan, "Leveraging Linguistic Context in Dyadic Interactions to Improve Automatic Speech Recognition for Children," *Comput. Speech Lang.*, vol. 63, Sep. 2020, doi: 10.1016/j.csl.2020.101101.

[14] I. Medhi Thies, N. Menon, S. Magapu, M. Subramony, and J. O'Neill, "How do you want your chatbot? An exploratory Wizard-of-Oz study with young, Urban Indians," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10513 LNCS, pp. 441–459, doi: 10.1007/978-3-319-67744-6_28.

[15] J. Chubb, S. Missaoui, S. Concannon, L. Maloney, and J. A. Walker, "Interactive storytelling for children: A case-study of

design and development considerations for ethical conversational AI," *Int. J. Child-Computer Interact.*, no. xxxx, p. 100403, 2021, doi: 10.1016/j.ijcci.2021.100403.
[16] V. Objects, "Myscript handwriting recognition engine." 2009.
[17] D. Ha and D. Eck, "A neural representation of sketch drawings," *arXiv Prepr. arXiv1704.03477*, 2017.