Cocomix: Utilizing Comments to Improve Non-Visual Webtoon Accessibility

Mina Huh
School of Computing, KAIST
Daejeon, Republic of Korea
minarainbow@kaist.ac.kr

Yunjung Lee
Computer Science and Engineering,
Ewha Womans University
Seoul, Republic of Korea
yunjung_lee20@ewhain.net

Dasom Choi
Department of Industrial Design,
KAIST
Daejeon, Republic of Korea
dasomchoi@kaist.ac.kr

Haesoo Kim
School of Computing, KAIST
Daejeon, Republic of Korea
haesookim@kaist.ac.kr

Uran Oh
Computer Science and Engineering,
Ewha Womans University
Seoul, Republic of Korea
uran.oh@ewha.ac.kr

Juho Kim
School of Computing, KAIST
Daejeon, Republic of Korea
juhokim@kaist.ac.kr
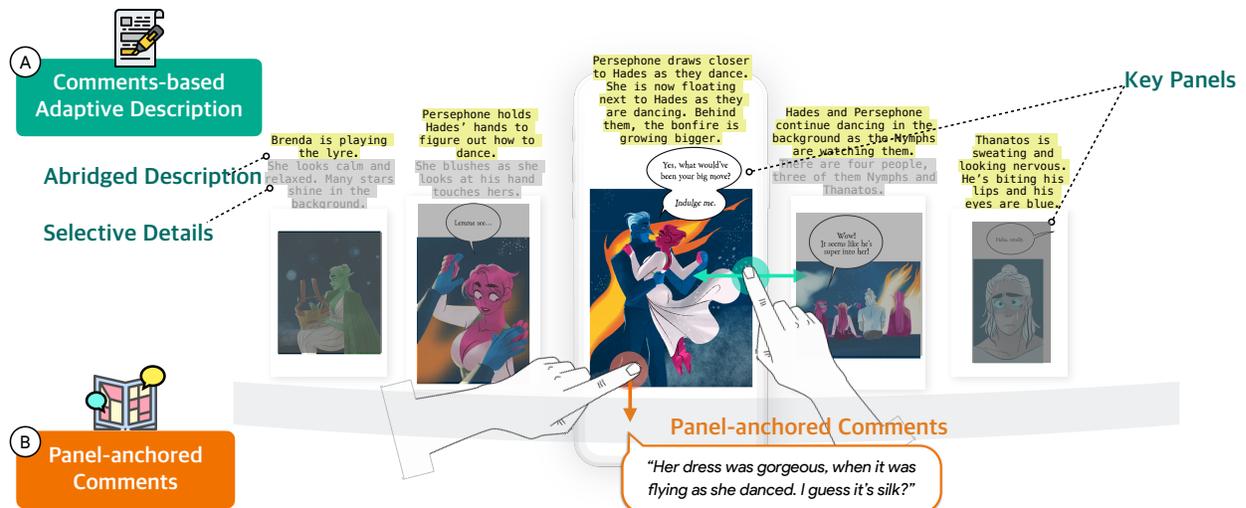
Figure 1: Cocomix is an interactive webtoon reader that leverages comments data to provide non-visual access to webtoon contents. Cocomix has two main components: (A) comments-based adaptive descriptions and (B) panel-anchored comments. With (A), readers can listen to more detailed descriptions in *key panels* which are frequently mentioned in comments. Abridged descriptions are provided for remaining panels as default and readers can selectively access more details. With (B), readers can easily access descriptive comments relevant to the current panel. Image from [WEBTOON SERIES] (2022).

ABSTRACT

Webtoon is a type of digital comics read online where readers can leave comments to share their thoughts on the story. While it has experienced a surge in popularity internationally, people with visual impairments cannot enjoy webtoon with the lack of an accessible format. While traditional image description practices can be adopted, resulting descriptions cannot preserve webtoons' unique values such as control over the reading pace and social engagement through comments. To improve the webtoon reading experience for BLV users, we propose *Cocomix*, an interactive webtoon reader that leverages comments into the design of novel webtoon interactions. Since comments can identify story highlights and provide additional context, we designed a system that provides 1) comments-based adaptive descriptions with selective access to details and 2) panel-anchored comments for easy access to relevant descriptive comments. Our evaluation (N=12) showed that Cocomix users could adapt the description for various needs and better utilize comments.

CCS CONCEPTS

• **Human-centered computing → Accessibility systems and tools**; **Empirical studies in accessibility**.

KEYWORDS

Accessibility, Digital comics, Webtoons, Image description, Comments

ACM Reference Format:

1 INTRODUCTION

Webtoon, a portmanteau of "web" and "cartoons," is a type of digital comics published and read online. Optimized for web and mobile consumption, webtoons abandoned the traditional layout of panels and adopted the single-strip display where readers navigate panels with scrolling or swiping gestures. Unlike traditional media channels for comics, webtoon platforms facilitate active participation of readers by enabling them to add comments on each episode. Also, the short reading time per episode has enabled readers to enjoy webtoons as a lightweight medium, often referred to as 'snack culture' [1]. While webtoons originated in South Korea, there has been a surge in popularity internationally. For instance, Naver Webtoon [2] [3] has more than 67 million monthly active users (MAU) worldwide. Together with the commercial success of adaptations into films or dramas (e.g., *Sweet Home, Along with the Gods*), webtoons now became one of the mainstream cultures in many countries.

However, blind or low vision (BLV) people cannot enjoy this popular medium because of the accessibility barriers posed by the foundation of comics—combining graphics and text. Incapable of participating in what is considered the cultural mainstream, BLV people feel excluded [74]. The most standard and economical way of providing accessible version of images is via textual description. However, it is yet to be explored how the conventional image description practices can be generalized to webtoons, a unique media as a sequence of images.

To better understand BLV readers' needs in webtoon descriptions, we conducted formative interviews followed by co-reading exercises with 10 BLV participants. From the two-stage formative study, we have identified the following challenges when adopting traditional description practices in webtoons. First, the flow of the plot is disturbed by the repetitive information in image descriptions. Second, BLV readers have difficulty in processing the given information when reading long descriptions. Third, BLV readers lack control over the amount of details in the description. Finally, BLV readers have difficulty in finding comments that are useful for understanding the content.

To overcome the challenges, we leverage comments data to improve webtoon descriptions. We believe that comments can be utilized to enrich webtoon descriptions, as they implicitly highlight key moments in the content through the reactions of readers, and provide additional context to the story. While comments data present a unique opportunity in augmenting visual descriptions for BLV people, to our knowledge, researchers have not yet attempted to leverage comments as the meta data for accessibility. To explore the feasibility and limitations of using webtoon comments for eliciting useful information for BLV readers, we performed a qualitative content-analysis of a set of 1,000 comments. From the results, we distilled the criteria of *descriptive comment*, which informed our approach of extraction and usage of comments.

We developed *Cocomix* (Figure 1), a prototype webtoon reader that provides efficient and interactive webtoon descriptions to BLV users through comments-driven interaction techniques. Our computational pipeline automatically calculates the importance of each panel based on the comment data. BLV users can listen to the description with adaptive levels of detail depending on its importance (Figure 1.A). To provide additional context, Cocomix also extracts descriptive comments and links them to corresponding panels (Figure 1.B). While reading, BLV users have active control over the reading pace and the level of detail by requesting additional details or relevant comments per panel on demand.

To assess the feasibility of Cocomix for enhancing the webtoon experience for BLV readers, we conducted a within-subjects study with 12 BLV people. Participants found Cocomix's comments-driven interactions useful for controlling the reading pace and utilizing comments. They also enjoyed having options to adapt the description for different reading contexts such as light reading and detail-focused reading.

This paper makes the following main contributions:

- Identification of high-level goals for webtoon accessibility
- A design approach that leverages comments data to improve the webtoon reading experience, backed up with the analysis of 1,000 webtoon comments
- Cocomix, a prototype webtoon reader that supports an interactive webtoon reading experience
- An empirical evaluation showing the value of comments-driven interaction techniques

2 RELATED WORK

We review previous research in accessible representations of images (Section 2.1), background for understanding comics, webtoons and research around them (Section 2.2), existing approaches for accessible comics (Section 2.3), and prior works on leveraging comments as a supplemental knowledge source (Section 2.4).

2.1 Accessible Image Representations

Digital imagery is pervasive in the online world, facilitating the sharing of information and ideas. BLV people often use screen readers and screen magnifiers [47, 74, 78] to access them. In order for a screen reader to render an image, an alternative text description [72] should be provided which can be generated manually by humans or automatically by using object recognition techniques [79]. With the advancement in computer vision techniques, prior research have increased the image accessibility at scale [27, 79]. For interactive

---

[1]https://en.wikipedia.org/wiki/Snack_culture

[2]https://www.webtoons.com/en/

[3] This work is not endorsed by WEBTOON. Views, opinions, and thoughts in this work are solely of the author.

and personalized alt-text, Morris et al. [48] has proposed alt-text with progressive-detail and question & answer.

A number of studies have been conducted to add image descriptions to various classes of images such as STEM images [46], memes [26], and GIFs [25]. For eyes-free art, Rector et al. [60] proposed proxemic audio interfaces. For artwork exploration using touch screen, Ahmetovic et al. [1] proposed attribute-based and hierarchical exploration. Also, EyeDescribe [61] utilized eye gaze data and a spoken description for image labelling.

There have been many guideline-level suggestions for representations of various classes of images. In this paper, we focus on a class of images which remains relatively inaccessible — webtoons.

2.2 Comics as a Unique Medium

Broadly, comics are "Juxtaposed pictorial and other images in deliberate sequence, intended to convey information and/or to produce an aesthetic response" [41]. Traditional comics are printed on papers, where each page is composed of a set of *panels* presenting a specific situation. The panels are often separated by a white space area named *gutter* [11]. Because of this gutter, all comics have the concept of *closure*, the "phenomenon of observing the parts but perceiving the whole" [41]. With closure, readers can comprehend the meaning between neighboring panels.

After a slump in the traditional comic industry, comics turned their stage to digital platforms to bring a broader audience. The infiltration of comics into the digital world has not only challenged the traditional structure of comics but also enabled anyone to publish and access comics [34]. Webtoon is a type of digital comics that adopts a vertical layout so that users can scroll down on web platforms and mobile devices. The amount of materials published in webtoon form has reached an equal amount as that published offline[4]. While there are other types of web comics published in online platforms or social media such as Tumblr, Tapastic, Facebook, and Instagram, we focus on webtoons based on their large readership and the relatively structured and consistent format.

Research around comics has been conducted in contents analysis [11, 15, 52, 53], content adaptation [6, 37], or interactions [59, 70, 76, 77]. Iyyer et al. [29] explored how computers can understand the closure-driven narratives in comics. To address the loss of author-intended segmentation in digital comics, Wang et al. [76, 77] introduced the concept of *phasels*, a new segmentation of panels that describes a strong relation among certain panels. Our work builds upon prior research on comics interactions by exploring non-visual interaction techniques while focusing on webtoons.

2.3 Making Comics Accessible

To make an accessible version of comics, tactile books have been designed. For example, braille formats [12, 14, 23, 51, 69] and textured images [58] were proposed so that BLV people can read comics by touch. Other approaches have investigated how to improve accessibility in digital platforms by translating the visual content to speech and sound effects. For instance, GraphicAudio [10] and Unseen [3] presents audiobooks that can be read without an aid of a screen reader. Rayar et al. [59] devised a digital comic viewer with magnifying features for people with low vision, and Lee et al. [36]

explored BLV people's preferences on audiobook and eBook format of digital comics. To create comic descriptions, Samson et al. [63] explored the potential of fan readers in describing comics. NNELS [55] has worked on developing a standardized approach to creating a "Described Comic Book", yet, to the best of our knowledge, no prior work has attempted to create descriptions specifically for webtoons. Rayar et. al. [58] argued that desirable solution approaches should be less expensive, less time-consuming and allowing access to an important part of the culture. Inspired by this, we explore how to support BLV population to be part of the webtoon culture.

2.4 Comments as a Supplemental Knowledge Source

Comments are a type of user engagement to express opinion or feelings visible to platform users [2]. Comments are capable of adding supplemental information to the original content, which can be interesting to other users [45]. To understand the commenting and rating behavior, Siersdorfer et al. [64] analyzed dependencies between comments and other metadata of the content. To understand how people contextualize comments, Yarmand et al. [81] explored reference patterns in YouTube comments. Also, Momeni et al. [44, 45] analyzed properties of *useful* comments for annotation of social media objects.

While reader comments contain diverse topics, conventional ranking mechanisms for comments (e.g., by posted date or by likes fail to provide an overview of topics discussed in comments [40]. For efficient browsing of comments, Opinion Space [19] and StarryThoughts [33] support navigation of diverse comments through visualization. RevMiner investigated extractive presentation of attribute-value pairs to summarize unstructured texts.

There have been suggestions for utilizing online reviews in the accessibility domain [38, 67]. Revamp [75] supports interactive information retrieval by leveraging customer reviews. For comments, Voykinska et al. [74] has observed that BLV people rely on others' comments to understand visual contents in social networking services. Still, to our knowledge, no prior work has attempted to leverage comments to augment visual description. We look into the opportunities and challenges of using comments, which are inherently more personal and diverse in context than reviews.

Our work builds upon prior work on comments mining by analyzing patterns of webtoon comments, and designing comments-driven interaction techniques for webtoon accessibility.

3 FORMATIVE STUDY

We conducted semi-structured interviews with 10 BLV people followed by a webtoon co-reading exercise to better understand their needs in webtoon descriptions. We first present methods and results of the study. Then, we discuss challenges and design goals, and present guidelines for webtoon descriptions reflecting the findings.

3.1 Methods

3.1.1 Participants. We recruited 10 BLV people (6 male, 4 female) interested in webtoons through mailing lists (Table 8). They averaged 26 years old (min = 22, max = 35) and described their vision as either blind (7 participants) or low vision (3 participants). All

---

[4]http://everything.explained.today/Webtoon/

Webtoons ID	Genre	Title (episode)	Length	Drawing Style	Popularity (likes)	URL
W1	Sports	The Boxer (1)	long	reality	99K+	[31]
W2	Slice of life	Independent Diary (77)	short	abstract	99K+	[30]

Table 1: Two webtoons used in the formative study for the co-reading exercise

participants were screen reader users. Three participants had prior experience with paper comics before losing sight. We provided a 15,000 KRW (approx. $12 USD) compensation for their time.

3.1.2 Semi-structured Interview. In the semi-structured interviews, participants were asked about their motivations for reading webtoons, prior experiences with webtoons, their perceived accessibility, strategies to access webtoons, and accessibility barriers that they encountered with those approaches.

3.1.3 Co-reading Exercise. We also conducted webtoon co-reading exercises to see what kind of information is desired in webtoon descriptions. BLV participants read two webtoons through a default image description provided by researchers. After one researcher read the default image description of each panel out loud, participants were encouraged to ask for further details they wish to know from the panel. We selected two webtoons (Table 1) from Naver's most-popular-by-genre [5] page that represents different genres and drawing styles.

In the default image description, we provided a brief visual description of scene (story setting, objects, characters, actions) per panel following established guidelines of [55]. All textual content such as dialogue and onomatopoeia were delivered. The reason behind providing a minimum default description was to encourage BLV participants to actively request additional information so that we can analyze types of information desired in webtoon description.

3.1.4 Analysis. Each session of interview and co-reading exercise was audio-recorded, transcribed [6], and coded by three of the researchers following the thematic analysis methods [24]. To analyze the types of information requested in the co-reading exercise, two of the researchers labeled and divided questions into three groups based on the underlying intent: visual information seeking, opinion seeking, and confirmation seeking. The distinction between information seeking and opinion seeking questions was based on whether the question asks about factual information such as "What is the color of his shirt?" or asks about relatively subjective information that involves researchers' opinion like *"Do you think the guy is good-looking?"*. Questions were categorized as confirmation seeking if they were yes-no question about the correctness of participants' interpretation's such as *"So the character is still in the gym, right?"* or if they asked researchers to repeat the description.

3.2 Results

In this section, we present BLV people's motivation for reading webtoons (Section 3.2.1), prior experiences with webtoons (Section 3.2.2), and the desired information in webtoon descriptions (Section 3.2.3). Then, we present challenges and design goals to improve webtoon reading experiences for BLV readers.

3.2.1 Motivation for reading webtoons. All participants answered that they wish to be able to read webtoons. In particular, seven participants specified that they wish to get involved in what is considered the 'mainstream'. They frequently encountered webcomics posted in social media (P1, P3, P7) or movies and drama adaptations of webtoon (P1, P4, P7), and wished to be able to join in on the conversation elicited from these particular forms of media. Six of the participants (P1-P3, P7, P8) said that they believe they can have a better understanding of context in conversations with other people by reading webtoons. Participants noted that their friends often talked about how they enjoyed an episode (P5, P8), or sometimes referred to a webtoon character as a metaphor (P8). Four of the participants mentioned the unique characteristics of webtoons as the motivation for reading webtoons. They found the visual storytelling technique interesting (P5, P6), and admired the quick speed of reading (P3, P4, P5).

3.2.2 Prior experiences with webtoons. We asked participants about their prior experiences encountering webtoons on the web as a whole. None of the participants had encountered webtoons with an embedded alt-text. For workarounds, they asked friends or family members to verbally describe the content (P1, P3, P7, P10), watched an adaptation such as webtoon-based dramas or movies (P3), searched for summaries in the fandom wiki (P6), or used image recognition apps (P1,P6). Still, participants mentioned the limitations of these approaches. Asking others for help was not always viable and referring to adaptations or summaries did not cover some information that they could only gain from reading the original material. Also, image recognition apps were not reliable due to low accuracy. Six of the participants mentioned that they refer to comments posted by other readers for each episode to guess the content (P1, P3, P5, P6, P9, P10). However, they reported that it is difficult to utilize comments as many of them are not directly related to the story and thus not helpful in understanding the content.

3.2.3 Desired information in webtoon descriptions. In the co-reading exercise, a total of 307 questions (117 asked in W1, 190 in W2) were asked by the participants. More than half of the questions (62%) consisted of participants seeking visual information and among them, only 1% of the questions were about drawing styles or view angle and none of the participants asked about the panel layout. Twenty percent were opinion seeking questions and eighteen percent of the questions were confirmation seeking questions. We observed that this type of questions were frequent especially when reading panels near the end, as readers had to keep track of all the information received up to that point. Also, participants who asked many questions about visual information requested for confirmation more often compared to others, as they had to process more information.

---

[5]https://www.webtoons.com/en/top
[6]https://clovanote.naver.com

3.3 Challenges and Design Goals

Combining both interview responses and observations from co-reading exercise, we identified four challenges that BLV people face when reading webtoons via image descriptions:

C1. Confusion in understanding the plot with recurring details in image descriptions
C2. Difficulty in processing the given information when reading the long description
C3. Lack of control over the amount of details in the description
C4. Difficulty in finding comments that are useful for understanding the content

The first challenge (C1) is the issue of repetitive information interrupting the flow of the plot. During the co-reading exercise, P2 reported that "Sometimes, same details are repeated, like the character's pose even when it hasn't changed. It disturbs the reading process." While webtoons consist of a sequence of images, the individual images come together to form a comprehensive plot [41]. This means that following traditional image description practices would not be appropriate for webtoons. We recognized that there would be a need to create a separate, webtoon-specific set of description guidelines to minimize information redundancy and would be more optimized for conveying the plot.

The second challenge (C2) is that BLV readers find it difficult to process all given information when listening to long descriptions. Compared to image posts on social media where few alt-texts can cover the content, webtoon descriptions tend to be much longer. Also, unlike a described video where audio cues help in picturing the scene, the delivery of a webtoon's content has to rely solely on written description. When given a long-winded description, BLV readers often face difficulty in processing and remembering the information provided. P7 mentioned "Whoa, can you repeat everything again? There is just too much information to handle. Descriptions, dialogues, and for dozens of images!". This becomes especially challenging when readers are not familiar with the setting or characters. For efficient delivery of information, three participants (P2, P5, P10) suggested differentiating the amount of details depending on the importance of elements. For example, when P5 was presented with an image of multiple characters, he asked for each character "Do you think this person is important? If not, can I skip? I'm not interested in his hair color if he won't show up again."

The third challenge (C3) is that BLV readers have no control over how much detail is provided in the description. From the co-reading exercise, we observed that the number of questions asked by participants varied considerably, which implies that every BLV reader has a different taste in description. However in conventional image descriptions, the amount of information delivered is solely decided by describers. This takes away one of the unique values of reading webtoons, the active control over reading pace.

The final challenge (C4) is that BLV readers have no efficient ways to cherry-pick comments that are helpful for understanding the content. While referring to comments was one of the main approaches to guess the content or complement the description, all participants who attempted to read comments reported the inconvenience of utilizing them. For example, P1 mentioned "Most comments are like "This is hilarious!" and I am left wondering what
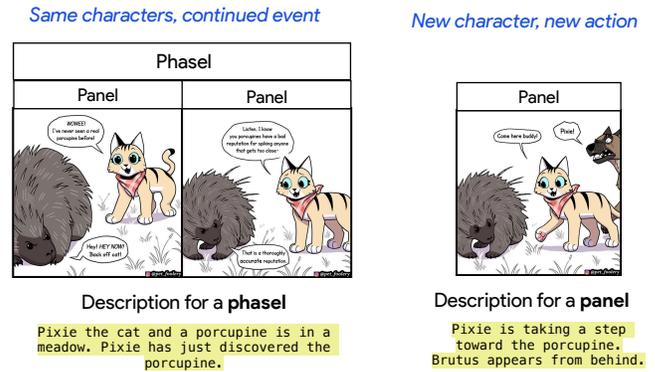
Figure 2: A phasel is a set of multiple panels that can be grouped or summarized into a single unit. Our guidelines suggest grouping similar panels, such as when the same people are appearing or a single action is described over multiple panels, into a single phasel to avoid repetition. The grouping can only happen between neighboring panels. Image from [WEBTOON SERIES] (2022).

is hilarious." From the feedback, we distilled the need for extracting comments that are useful for understanding the content.

Based on the challenges, we identified four design goals for systems to improve webtoon reading experience for BLV people. Two of the design goals address how the description can be more efficient (D1, D2), and the other two design goals address how users can interactively control the reading (D3, D4).

D1. Provide descriptions with minimized information redundancy (C1).
D2. Provide adaptive levels of detail for each panel based on readers' collective interest (C2).
D3. Present additional details on demand to support active control over the reading pace (C2, C3).
D4. Support selective access to comments that describe the content (C4).

3.4 Guidelines for Webtoon Descriptions

Based on the findings from our formative study, we also discovered a need for an improved methodology of describing webtoons. We aimed to improve upon existing comic description protocols by focusing on the unique aspects of the webtoons, as well as focusing on the user needs reported while reading webtoons.

3.4.1 Methods. To create the guideline, four researchers individually created descriptions for a single webtoon. After collecting the descriptions, we grouped common themes and strategies used and discussed which strategies were most successful in fulfilling the challenges, as well as what might be potential ambiguities that might arise due to difference in comprehension. Then, we developed comprehensive guidelines that detailed how and what we should be described. The structure of our guideline was followed other guidelines for image accessibility [8, 17].

Genre	Title (episode)	Length	Drawing Style	Popularity (likes)	URL
Drama & Romance	True Beauty (140, 141)	long	reality	6.4M	[80]
	The Remarried Empress (72, 73)	long	reality	1.9M	[4]
Fantasy	Lore Olympus (162, 163)	long	semi-abstract	5.2M	[65]
	Familiar Feelings (11-1, 11-2)	medium	semi-reality	293.8K	[32]
Comedy	ZomCom (138, 139)	short	semi-abstract	724.7K	[18]
	Pixie and Brutus (94, 95)	short	reality	730.2K	[21]
Action	unOrdinary (223, 224)	long	semi-reality	5.1M	[73]
	Hybrid (29, 30)	medium	semi-reality	114.1K	[9]
Slice of life	Murrz (510, 511)	short	semi-abstract	773.6K	[49]
	Safely Endangered (672, 673)	short	abstract	1M	[42]

Table 2: 10 Webtoons used in the comments analysis

Category	Sub-category	Definition	Example Comment
Description (42.2%)	Visual description (12.7%)	Visually describe a specific part (e.g.. appearance, color, shape)	"John's power towards the end with all of the black tentacles and sickly colors is like the physical representation of trauma"
	Conceptual description (15%)	Conceptually describe a specific part (e.g., personality or emotion of a character)	"I can see that Brutus is nervous."
	Description of readers' emotion (14.5%)	Describe what readers felt with a specific part of the content	"Her glowing eyes sent chills down my spine!"
Non-description (57.8%)	Question (4.9%)	Ask for clarification for misunderstood part	"Wait, did the knife hit him or did he step on it or something?"
	Opinion on upcoming plot (7.9%)	Speculate about or wish for what will happen next in the story	"Once they know why Navier said yes, I bet Kosair will be 100% behind Heinrey succeeding"
	Communication (8.9%)	Communicate to the creator or other readers	"Happy Early Birthday Author"
	Self-expression (7.6%)	Share one's experience or opinion outside of the story	"I'm hoping me seeing this means I will get the interview for the city job I just applied to!"
	Assessment of content (4.3%)	Assess the quality of the work	"Can we take a moment to talk about how amazing the plot is?"
	Others (24.2%)	Inconclusive comments with minimum context (e.g., puns, emoji stuffings, etc)	"Call the FBI. The rice has gone rogue"

Table 3: Taxonomy of descriptiveness with example comments.

3.4.2 Contents of Guidelines. The guideline (in supplementary material) consists of three main themes: 1) how to describe metadata not directly related to the plot, such as title. 2) how to subdivide a webtoon into, and write descriptions for, scenes and *phasels*, and 3) writing styles such as how to express speech and level of detail in image descriptions. In the second theme in particular, we borrow and adapt Wang et al.'s concept of phasels to ensure non-redundant descriptions of elements that span multiple panels [76, 77]. In this paradigm, panels could be combined into a single phasel (Figure 2), reducing the need for redundant explanations and user interactions (C1). In later sections, we describe how descriptions following our guidelines are used as input data for our system.

4 WEBTOON COMMENTS ANALYSIS

From the formative study, we discovered that participants often refer to comments to infer webtoons' visual content, yet have difficulty in utilizing them. To further explore the potential of webtoon comments as the meta data for accessibility, we conducted comments analysis. This section presents our approach to explore the feasibility and limitations of using webtoon comments for eliciting useful information for BLV readers. In particular, we analyzed what elements the comments describe and how they are described.

4.1 Methods

4.1.1 Data collection. First, we crawled 1K webtoon comments (Table 2) such that the dataset would contain diverse representations of genre, length, popularity (by the number of *likes*), and drawing style – from realistic to abstract in McCloud's pyramid of the comic universe [41]. We sampled 20 webtoon episodes (5 genres × 2 webtoons × 2 episodes) from Naver [7], which operates the world's largest webtoon platform. For each webtoon, we chose their two latest episodes and collected 50 most-liked comments to simulate the default comment order set by the platform (20 episodes × 50 comments per episode). By only collecting top-level (no threaded) comments, we ensured that each comment is not dependent to another comment.

4.1.2 Analysis. We followed the thematic analysis methods [24]. As a first step, the lead researcher open-coded 100 comments and extracted initial themes. Then, two other researchers reviewed the same data, and discussed further for conflict resolution and validation. All researchers read all webtoons before analysis to reach an agreement on the interpretation. With another 100 comments, all three researchers iteratively developed the themes. Based on the themes, we derived a taxonomy for determining whether a comment is a *Description* or a *Non-description* (Table 3). Additionally, we developed a taxonomy for categorizing the *Description Target* and *Description Range* (Table 4).

4.2 Results

4.2.1 What makes a comment descriptive or non-descriptive? Reflecting the information needs of BLV readers in webtoon reading identified from our formative study, we set a broader range of *'descriptiveness'* than a mere visual description. We added the conceptual description and description of readers' emotion, which are not directly related to visual attributes of webtoons. Still, as Wang et al. [75] noted, these high-level attributes inferred from the visuals can lead to meaningful interpretations. Comments were categorized as *Non-description* (Table 3) when they included self-expression, communication, assessment of the content, or opinions on upcoming plots, or questions.

4.2.2 What do comments describe? As depicted in Table 4, we found that people describe characters more often than other elements of webtoons. For description range, most comments (46.1%) were describing a single panel. Comments categorized as multi-panel descriptions either referred to continuous panels with a single event, or contained several sentences each of which referred to a different panel. We found that descriptive comments were not evenly distributed in all webtoon elements. For example, only few characters were mentioned frequently, and only a small number of panels (min = 2 out of 4, max = 8 out of 46) were referred to in comments. While we did not analyze the intent behind each description, we speculated that people often describe a part of the story to deliver how impressive it was (e.g., *Her dress was gorgeous, especially when it was flying!*), or to call attention to a notable part that other readers are likely to miss (e.g., *The color of his eyes slightly changed during the dance! Did anyone notice?*).

---

[7]https://www.webtoons.com/en/

Category	Sub-category
Description Target	Character (55.7%)
	Object (7.4%)
	Dialogue (10.8%)
	Event (19.3%)
	Setting (2.7%)
	Others (1.6%)
	Inconclusive (2.5%)

(a) Target specificity

Category	Sub-category
Description Range	Single-panel (46.1%)
	Multi-panel (28.6%)
	Entire episode (7.3%)
	Inconclusive (18.0%)

(b) Range specificity

Table 4: Taxonomy of description specificity

4.2.3 What makes it difficult to utilize comments? Despite the abundance of descriptive comments, there exist barriers to fully utilize comments. First, comments are inherently unstructured. Sentences that are incomplete or conveying meanings with emojis have the risk of misunderstanding [71]. Second, echoing the C4 distilled from the formative study, it is difficult to locate descriptive sentences as the majority of comments are non-descriptive, or lack a clear reference target. With no filtering feature supported by the platform, BLV readers go through many non-descriptive sentences to find a useful comment. Third, the presentation order of comments is not helpful for utilizing comments as descriptions. Most platforms only provide the sorting of comments by the number of likes or the time uploaded. Consequently, descriptive comments are not presented as the story order and users cannot selectively read comments relevant to the panel of their interest.

While webtoon comments have the potential to add rich descriptions, we believe that efficient filtering and sorting would support better utilization of comments.

5 COCOMIX

With four design goals (Section 3.3) in mind, we present Cocomix, a prototype webtoon reader supporting efficient and interactive webtoon reading for BLV users through comments-driven interaction techniques. Below, we walk through a scenario illustrating how BLV readers can use Cocomix for reading webtoons (Section 5.1) and subsequently describe how the system features are enabled using comments (Section 5.2-5.3). We then also describe the computational pipeline that powers Cocomix (Section 5.4).

5.1 User Scenarios

5.1.1 Scenario 1. Sean is a college student with low vision who uses a screen reader to access web contents. Sean's friends often talk about how they enjoyed the new episode of a popular webtoon,
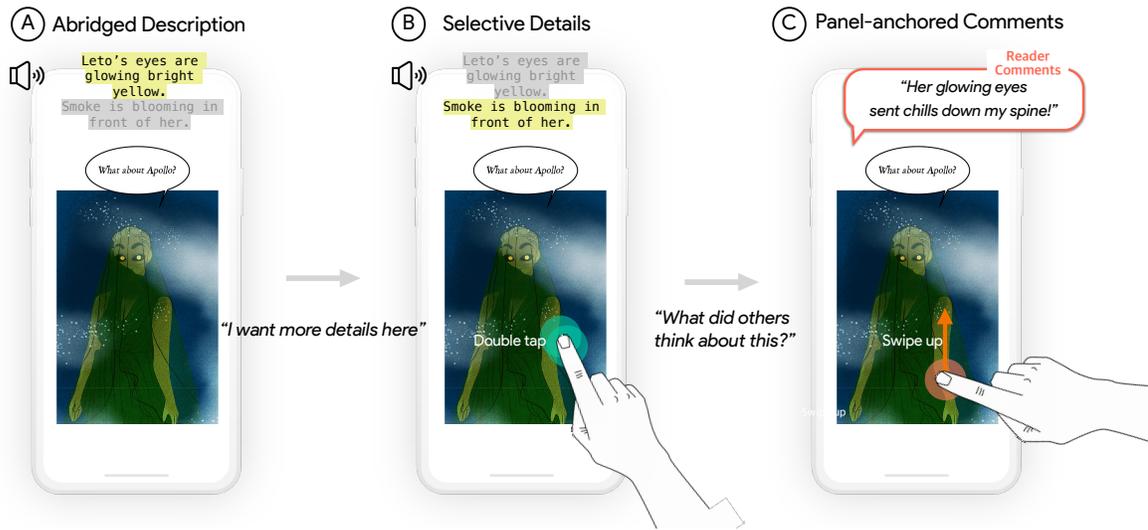
Figure 3: Readers can have personalized control over description using Cocomix. (A) For non-key panels, abridged description is provided as default. (B) When readers want to uncover more details, they can double-tap to listen to additional description. (C) When readers are curious about descriptive comments relevant to the current panel, they can swipe-up to access panel-anchored comments. Image from [WEBTOON SERIES] (2022).

"Lore Olympus", so he is interested in reading the webtoon. He wants to enjoy webtoons while commuting to school by subway, and he decides to use Cocomix, a mobile webtoon reader to read the webtoon. Unlike other audio comics that he has experienced before where the description was played continuously, Sean can navigate between panels by swiping the screen left-to-right. Because the gestures are similar to those of the screen reader he uses, he easily gets used to the system. As Sean progresses into a different scene, the system informs him of the transition of time and setting: "Nighttime, at the kitchen". In the story, the character Persephone is doing the dishes. As he moves on to the next panel, the system description tells him that Persephone has looked up from the dishes, looking confused, at her visitor. The visual difference between neighboring panels is clearly described, so he has no difficulty in connecting the panels to understand the whole story. Sean notices that some panels have shorter descriptions than others especially when the scene is focused on the dialogue, or when only walk-on characters appear in the panel. Because he wants to finish the episode in 10 minutes before arriving at school, he just listens to the abridged version of the description (Figure 3.A).

5.1.2 Scenario 2. After taking all classes, Sean comes home and decides to continue reading the next episode while relaxing. This time, he wants to slow down the pace of reading and peek into more details to picture the scene clearly in mind. While listening to the default description, Sean realizes that the main characters, Persephone and Hades, are described in more detail than other minor characters such as nymphs. This time he is curious about how the nymphs look so he double-taps the panel (Figure 3.B) where nymphs appear to request more details. From the audio feedback, Sean notices that some comments are anchored to specific panels.

While most comments are shown at the end of the episode, these anchored comments can be read while following the main story. When reading the scene where Persephone dances, he checks other readers' comments on the scene. By reading the panel-anchored comments (Figure 3.C), he could access extra visual information that was omitted in the description (e.g., "Some nymphs were holding hands while watching the couple dance. How cute!").

5.2 Adaptive Description with Selective Details

To address D2 and D3, Cocomix supports an adaptive description method based on the description generated through our guideline 3.4. The adaptive description provides an abridged description for panels that were mentioned in fewer comments, and a full description for panels mentioned in many comments (key panels). This is so that readers would not have to spend excessive time listening to details in less important panels. Through this approach, the average description length is reduced and readers can have less cognitive load while reading webtoons. Details on the abridged panels omitted by default can still be accessed on demand, which we call as *selective details* [48]. When reading with Cocomix, users are provided with audio feedback which informs them whether the panel has extra details to disclose or not, as to minimize the frustration when users face unavailable details or unknown-unknowns.

5.3 Panel-anchored Comments

To address D3 and D4, Cocomix provides panel-anchored comments where readers can refer to descriptive comments relevant to the current panel during reading. Cocomix first extracts comments that are descriptive or helpful for understanding the story then link them to the relevant panels. Comments that are not linked to

specific parts of the story yet still are descriptive are shown at the end of the episode.

5.4 Computational Pipeline

The computational pipeline that powers Cocomix has three main components: 1) scoring and abridging descriptions, 2) extraction of descriptive comments, and 3) linking comments to relevant panels (Figure 4). The first component enables adaptive descriptions while the other two components implement panel-anchored comments.

5.4.1 Data Collection and Preprocessing. There are two main data sources for our pipeline: 1) webtoon descriptions and 2) webtoon comments. After we collect top-level comments we discard ones that are 3-words-long or shorter. Then, we penalize low-quality comments using $like-score = number\_of\_likes - 2 * number\_of\_dislikes$ and dropping comments with a $like-score$ less than 10.

Because comments are inherently noisy and informal, we go through several preprocessing steps prior to applying Natural Language Processing (NLP) techniques. First, we remove emojis and emoticons as often the translated text does not align with the intent of usage. Next, we use Bing Spell Check API [8] for correction of misspelling and informal languages. Reflecting the observation from the comments analysis, we apply filtering and scoring techniques in sentence-level instead of comment-level, as often the sentences in a single multi-sentence comment refer to different panels. To improve the sentence-tokenization accuracy, we correct punctuation errors and casing errors by running transformer-based punctuator [43] and truecaser [68]. We run a language-detection algorithm to drop non-English comments [43] limiting our scope to producing outputs in English. Finally, we run co-reference resolution [82] before tokenizing each comment into sentences to recover from the loss of named entities when split.

5.4.2 Scoring and abridging descriptions. We now describe a method to detect key panels of webtoons. Ideally, key panels should capture the visual, conversational, or plot highlights, or have a significant role in understanding neighboring panels. However, current computer vision techniques are not advanced enough to automatically detect key panels in webtoons due to their high variety of visual styles and the scarcity of labeled data [11]. Thus, we rely on comments data to detect *"key panels"* based on our insight that an important panel is likely to be mentioned more frequently in the comments. Our pipeline first runs a part-of-speech tagger [28] to populate proper nouns, nouns, and verbs which correspond to characters, objects, and events with an intuition that people are likely to refer to those in comments. This reflects the findings from our comments analysis and the prior research in cognitive psychology [83]. We derive the focus score of a sentence S_i using scores of all pre-populated words, as we represent each sentence as $S_i = \{w_1, w_2, ..., w_k\}$. We define a sentence's focus score $Foc(S_i)$ as:

$$Foc(S_i) = \frac{1}{|S_i|} \cdot \sum_{w_k \in S_i} CF(w_k)/PF(w_k) \qquad (1)$$

$CF(w_k) = $ (# of comments that has w_k) / (total # of comments)

$PF(w_k) = $ (# of panels that has w_k) / (total # of panels)

where $CF(w_i)$ denotes *Comment Frequency* and $DF(w_i)$ denotes *Panel Frequency*. Similar to document frequency, both represent the number of comments or panels containing word w_i. Here we penalize with panel frequency of a word to avoid naively giving high scores to frequently appearing elements. In the equation, $|S_i|$ is the length of sentence S_i in number of words including stopwords. A panel's focus score is determined by summing all sub-descriptions in the panel. We observed that the score computed is dependant on the type of webtoon, or other external factors—for example, when the published date was around the time of the content creator's birthday, most comments were for creators. Thus, instead of setting a threshold score to determine key panels, we empirically set top 30% scored panels as key panels with which researchers could perceive the effect of abridging while still retaining the key information.

For abridging the remaining panels, each of the panel descriptions is shortened to 30~50% of the original length. We do not run the abridging technique with dialogues (texts wrapped in speech bubbles) as a gap in the conversation can be critical in following the story. We apply extractive summarization using fine-tuned BERT [39]. Compared to abstract summarization which is similar to the process of paraphrasing and more prone to factual inconsistency [50], extractive summarization keeps the original structure. Sentences not included in the resulting summary can be accessed on demand (with a double-tap).

5.4.3 Extraction of descriptive comments. Reflecting the findings from the formative study and comments analysis, we extract comments that meet all of the following criteria: 1) have clear reference to an element of the webtoon, 2) be a form of statement, and 3) describe what happened rather than the upcoming plot (Figure 5). First, we run a part-of-speech tagger [28] to the webtoon description to pre-populate proper-nouns, nouns, and verbs (*webtoon keywords*) which correspond to characters, objects and actions. Then, only the sentences from comments data with exact word-level match with at least one of webtoon keywords are extracted. Also, comments that quote a dialogue with a continuous overlap with more than three words are extracted. The intuition is that descriptive comments would have a clear reference to an element of the story. Then we run a dialogue act classifier [57] which outputs tags that summarize syntactic, semantic, and pragmatic information. We only extract sentences with the tags 'Statement-non-opinion' (e.g., *Leto was hiding in the middle of the crowd!* or 'Statement-opinion' (e.g., *Thanatos' reaction was priceless, sweating bullets*) among 43 different tags. As a final step, we check the tense of a root word of a sentence and only keep sentences in past or present tense. This filters out comments with guesses (e.g., *Persephone will soon catch Leto's arms!* or hopes (e.g., *I will be so happy when they eventually get married!*) for upcoming stories as they are irrelevant to understanding the plot of the webtoon.

5.4.4 Linking comments to relevant panels. In Section 4, we observed that most descriptive comments consist of two parts: 1) making reference to a part of the story (e.g., *When Persephone was dancing [...]*) and 2) adding extra description (e.g., *[...] she looked so happy.*). Thus we assumed that descriptive comments would have high similarity with the relevant panel's description especially because of the first part. Semantic similarity between them
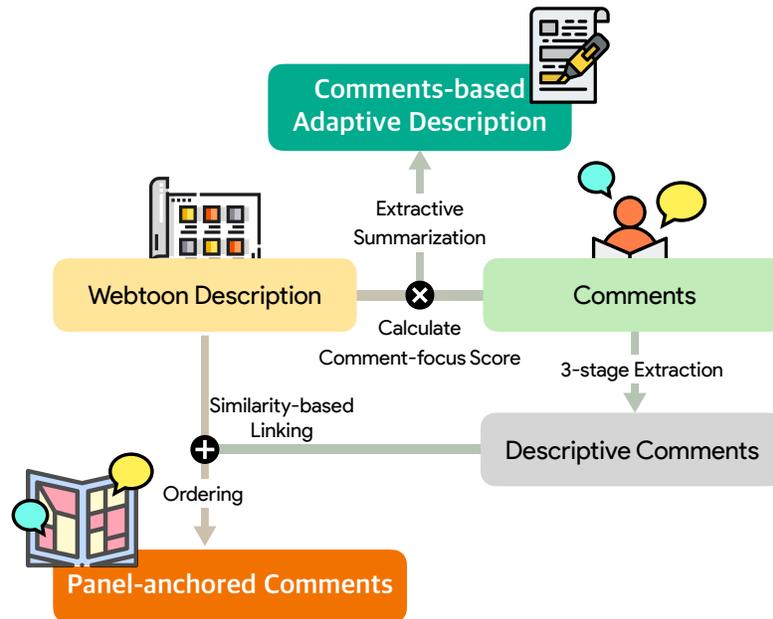
Figure 4: *Pipeline overview of Cocomix. It takes two inputs: webtoon description and comments.* **For comments-based adaptive description, our pipeline first calculates the comment-focus score to detect key panels. Then, to abridge the total length of description, we apply extractive summarization to non-key panels. The resulting description has different amount of details for each panel depending on its comment-focus score. For panel-anchored comments, our system first extracts descriptive comments via 3-stage extraction. Then, extracted comments are linked to relevant panels based on the similarity score and named entity match. Finally, comments are presented in an order that ensures non-redundancy.**

was computed using the sentence embeddings method of Arora et al. [7]. Here we use sentence-level linking instead of paragraph-level (entire comment or panel description) to keep the accuracy from dropping when there are multiple topics in a comment or a panel. Also, we check for named entities in the matched comment and drop if one not covered in the corresponding panel description appears. While the comment can be still relevant, we take the precision-first approach and filter out potential false-positives. For the decision of linking a comment, we check its similarity score distribution with all panels. The comment is linked only when the score is above a threshold in a single panel or neighboring panels. In this paper, we empirically set the threshold to produce an accuracy of 85%.

5.4.5 Presenting comments. Readers can read relevant comments anchored at the end of each panel description. When the comment is anchored to multiple neighboring panels, we provide it in the latest panel to avoid giving a spoiler. Also, when there are multiple comments anchored in a same panel, we order them so that neighboring comments have low similarity score. This is to avoid recurring of similar comments, reflecting the formative study findings. Comments extracted as descriptive but not linked are presented at the end of the episode. The presentation order is determined by the number of descriptive sentences in the comment and likes. While removed for the text analysis, we deliver emoji-descriptions while

reading comments for rich delivery. Still, we follow the recommendations from Tigwell et al. [71] by removing repeated emojis and skipping comments with emojis in the middle.

6 EVALUATION

To demonstrate the feasibility of our pipeline, we conducted pipeline evaluation study. Then, we conducted user evaluation study to evaluate the effectiveness of Cocomix's comments-driven interaction techniques.

6.1 Pipeline Evaluation

To evaluate the effectiveness of our comments filtering and linking techniques, we collected subjective ratings on Cocomix's output comments from 10 sighted webtoon readers.

6.1.1 Materials. We selected five different webtoons that span different attributes (length, number of comments, genre, and drawing style). To acquire descriptions for the selected webtoons to power the system, we recruited 10 people to describe the webtoons by following our guidelines (Section 3.4). Each webtoon was randomly assigned by two describers to ensure task completion, but they worked independently. We recruited external describers with no prior experience in description tasks to ensure the generalizibility of our pipeline. For episodic webtoons we selected the first episode to reduce the dependency on background context. For system input, one description was randomly selected in each webtoon.
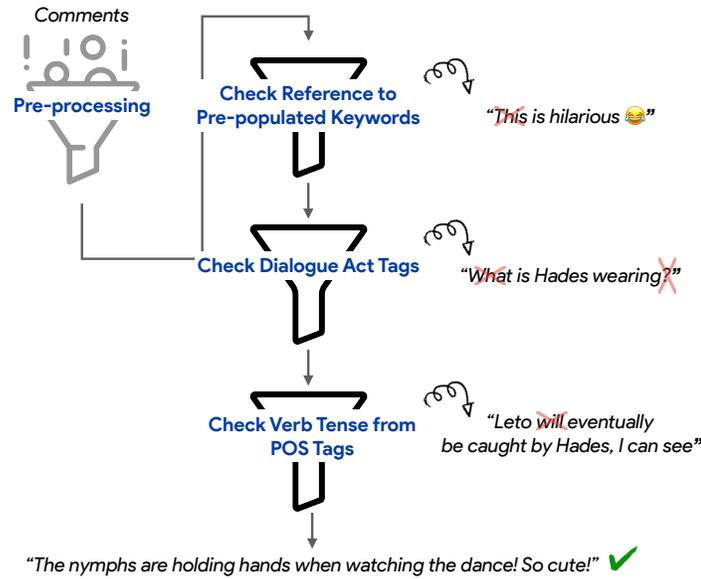
Figure 5: Descriptive comments are extracted through three stages of extraction. Prior to filtering, comments are pre-processed to auto-correct noisy and informal languages. Then, our pipeline first checks for reference to pre-populated webtoon keywords in each comment. This filters out comments with only non-clear reference. Second, it checks for dialogue act tags and leave out all comments except those with 'Statement-non-opinion" or 'Statement-opinion" tags. In this stage, comments in a question or response form are removed. Finally, the pipeline checks for the tense of the root word from Part-Of-Speech tags. This removes comments expressing wish or guess on the upcoming plot.

6.1.2 Procedure. We ran our pipeline with the five descriptions and comments, and distributed evaluation questionnaires to 10 evaluators (none of them participated in the description task). Here we recruited sighted evaluators to measure how well the output comments are in compliance with extraction criteria and liking criteria based on original visual content instead of descriptions. We also report BLV readers' evaluation of the comments (Section 6.2). To evaluate our extraction of descriptive comment, we randomly selected three from the pipeline-extracted comments and three from the 50 most-liked comments. For each comment, evaluators were asked to report how much they agree about the following: 1) The comment is related to the story, 2) The comment includes objective descriptions of (a part of) the story, 3) The comment includes subjective descriptions of (a part of) the story, 4) The comment refers to specific elements of the content (e.g., a scene, character, object, setting, mood). To evaluate our linking to relevant panels, we randomly selected one of our panel-anchored comments. Evaluators were presented with the panel-comment pair and asked to answer how much they agree with the following statements: 1) The comment is relevant to the panel; 2) The comment is targeting the panel.

6.1.3 Results. The results can be seen in the Table . While Cocomix-picked comments received higher scores than most-liked comments in every criteria, note that the scores were comparable for the coverage of subjective information (Table 5). Also, results about linking show that the accuracy of panel-anchored comments is acceptable (82% of the outputs were rated above 4 out of 7). Cocomix-picked

comments received low scores when they covered information about multiple panels but were matched to only one of them.

6.2 User Evaluation

To assess the feasibility of using comments-driven interaction techniques to enhance webtoon experience for BLV readers, we conducted a user study comparing the webtoon readers with and without our comments-driven interaction techniques. We explored three research questions:

(1) RQ1. How do BLV readers use Cocomix's interactive features while reading the webtoon description?
(2) RQ2. How do BLV readers perceive comments data presented in Cocomix?
(3) RQ3. How usable and useful are Cocomix's features?

6.2.1 Interfaces. The baseline interface stripped all comments-driven features from Cocomix, but still supported navigation between panels and unfiltered comments provided at the end of the episode. Both the system and the baseline provided description that followed our guideline as there is no established guidelines for webtoon description to compare with.

Gesture logs were collected using Google Firebase[9]. While our prototype works without an aid of a screen reader, we provide sufficient audio feedback like using a screen reader (e.g., when the user makes double taps, it gives a short "click-click" sound effect). In our study, we did not display images in order to explore how

---

[9]https://firebase.google.com/

Pipeline Component	Statements	Ratings for Comments	
		Most-liked	Cocomix
Extraction of descriptive comments	The comment is related to the story.	4.33 (1.78)	5.37 (1.45)
	The comment includes objective information of what happened in the story.	3.47 (1.71)	4.80 (1.54)
	The comment includes subjective information of what happened in the story.	4.67 (1.92)	4.89 (1.45)
	The comment refers to specific part of content.	3.43 (1.82)	5.23 (1.39)
Linking to relevant panels	The comment is relevant to the presented panel.	-	5.46 (1.79)
	The comment is targeting the presented panel.	-	5.02 (1.89)

Table 5: In the pipeline evaluation, we collected subjective ratings (in a 7-point Likert scale) on output comments of two pipeline components: Extraction of descriptive comments and Linking to relevant panels. The value represents the mean of scores (SD). Results show that Cocomix's output comments outperformed most-liked comments in describing the story.

well the visual content can be replaced with our description and comments-driven features. However, we agree that low vision users can benefit from images displayed, and wish to explore how features of Cocomix complement the image in our future work.

6.2.2 Participants. We recruited 12 BLV people (8 male, 4 female, Table 8) using mailing lists and social media. The criteria for invitation were interest in webtoon, using screen reader due to visual impairment, and basic English proficiency as the description was provided in English. They averaged 30.33 years old (min = 22, max = 53) and described their visual impairment as totally blind (9 participants) or low vision (3 participants). None of the participants has participated in the formative study. We compensated each participant $30 USD for their time.

6.2.3 Materials. We selected three webtoons from Naver (Table 6). We selected the first webtoon (W3) with short length to use in tutorial sessions. Webtoons used in main studies were selected from most-popular ongoing series, and were similar in terms of length (the number of panels), the amount of information provided in text (e.g., dialogue and descriptions), and the number of comments, but different in genre and creator. Following our guidelines (Section 3.4), two authors of this paper drafted the description for each webtoon and cross-checked for major omission, grammatical errors, and typos.

6.2.4 Procedure. A 90-minute study was conducted remotely using Zoom [10] for safety during COVID-19. Each study was audio-recorded and transcribed [11] for further analysis. The study was approved by our institution's IRB. After collecting informed consent, participants were given a tutorial of the Cocomix interface, and they were asked to go through a practice session using tutorial webtoon (W3) to familiarize with the features.

Then, a within-subjects study was conducted where participants were asked to read two webtoons. The presentation order of reader type (baseline vs. Cocomix) and webtoon type was counterbalanced. To ensure they pay attention to the content, we instructed participants that they will be asked to summarize the story after each condition. We provided summary of previous episodes before reading each episode to provide background context to the participants to minimize the effect of prior knowledge in study results.
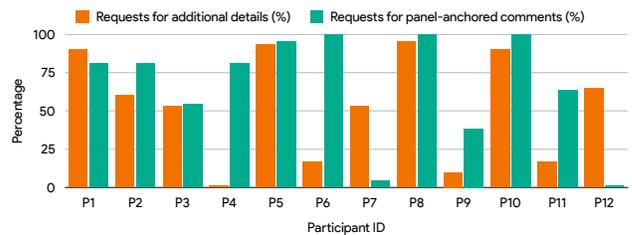
Figure 6: The percentage of panels where participants requested details or comments is shown for each participant. The usage pattern of Cocomix's interactive features varied by participant.

After each episode, participants completed a questionnaire on the usability of each interface and their experience and opinions about comments data. All gestural interactions were logged by the server for analysis. We ran semi-structured interviews at the end to gain a deeper understanding of the strengths and weaknesses of Cocomix and the participants' strategies for using interactive features.

6.3 Results

6.3.1 Usage Pattern of Interactive Features (RQ1). In both baseline and Cocomix conditions, all users sequentially followed the description order instead of going back or skipping a panel. Four of the participants (P11, P15, P18, P20) checked the details most of the time (Figure 6). P11 and P15 mentioned that they usually enjoy the slow pace in reading with lots of details. P18 and P20 explained the fear of missing out led them to request details every time. P18 noted *"I'm afraid of skipping any detail when I know (from the audio feedback) there is something. What if the system didn't read it to me and it later turned out to be important?"* Other participants (P14, P16, P19, P21) mostly skipped the details and read the default description. P19 mentioned *"The information given was enough for me! I was happy with that so just quickly moved on."*

While we couldn't observe a pattern from one participant, three participants showed a decrease in the number of requests for details as the reading continued. The reduced reliance on the details in the later parts of the story was due to getting familiar with the story context and building trust with the system. P17 commented

Webtoon ID	Title (# of panels)	# of Comments (Top-level / Total)	Genre	Creator	URL
W3	Pixie and Brutus (6)	521 / 744	Comedy	Pet Foolery	[22]
W4	Lore Olympus (72)	5,226 / 6,681	Romance	Rachel Smythe	[66]
W5	The Remarried Empress (64)	6,078 / 7,614	Fantasy	Alphatart & Sumpul	[5]

Table 6: Webtoons used in the evaluation study

"I requested details in the first few panels to see how important the hidden details would be. Then I soon realized they weren't necessary at all!" Also, P12 and P13 mentioned that they checked the details only when there were some comments left on the panel, as anchored-comments can be an indicator of notable points.

In the baseline, two participants (P19, P21) skipped the description and moved on to the next panel even when the audio was still being played. P19 reported *"After I experienced the short description in the first round (Cocomix condition), listening to the long one (baseline condition) was too boring. I didn't want to waste time."* He mentioned that he first listened to the few words of the description and when it was about minor character's appearance, he would skip.

6.3.2 Perception of Comments Data (RQ2). All participants were positive about reading other readers' comments. Participants mentioned the usefulness of comments in various aspects. First, P16, P18, P20, and P21 noted comments complement the description with additional details omitted in the original description. P18 noted *"I didn't realize that the bird's eye was purple before I read the comment! Describers are humans and they miss things. . . so thank you, commenters!"* P14, who never requested more details but actively checked most of the panel-anchored comments reported, *"I didn't even need to request more details as the comments gave details! And in a more fun way.* In addition, subjective comments were found to be helpful in picturing the scene more vividly. P18 noted *"The description said her eyes were glowing. The comment said 'Her glowing eyes sent chills down my spine!' I definitely prefer the comment's style, as it tells me how I'm supposed to feel it."*

Participants also noted that comments were useful even when they did not provide additional information about the story. They enjoyed comments as an independent component of a webtoon experience, where they could read humor, communicate, discuss the work and be part of the community and culture. P15 mentioned *"Even when I fully understood the story and didn't need further information, I just read comments because I'm curious about other readers' thoughts."* Also, P18 compared the two roles of comments and how the content should be different for each purpose. *"I enjoy some of the comments as a social activity after I finish an episode. Comments with this purpose can have more personal opinions, and it's totally fine. I feel like I'm connected to other readers and part of the conversation going around the story."*

One unexpected finding is that two participants (P15, P20) found comments useful for assurance of understanding the content of webtoons. P15 noted *"When I read other readers' comments and nothing unexpected is there, then I know whether I understood the scene correctly."* P20 also mentioned that *"When the comment adds additional details to what I heard from the description, it's good. But when the comment talks about completely new stuff, then (I*

experience) a dramatic loss of confidence" Similarly, three of the participants(P13, P15, P20) noted that they checked the details whenever comments were anchored to the panel. P13 mentioned *I can guess that something important is going on when many people are commenting about it (panel).*

6.3.3 Usability and Usefulness of Cocomix (RQ3). All participants reported Cocomix's interactive gestures (double-tapping to access selective details and swiping-up to access comments) were easy to learn. When asked which interface they wanted to continue using, 10 of 12 participants chose Cocomix (1 chose the baseline, 1 chose both). Most participants including participants who checked details every time (P15, P18, P20) mentioned that they preferred Cocomix as it provides more options in reading webtoons. Similarly, when asked if they had sufficient control over the amount of description, participants rated Cocomix significantly higher than the baseline (Figure 7). While the participants could still skip the description with the baseline, they mentioned it was difficult to interrupt the description is being played. P16 specified *"The amount of description accessible was the same in both conditions, but when the first version (baseline) was just dumping everything, second one (Cocomix) was more like asking "Would you like to know about this?' "*

Participants rated Cocomix's comments significantly more helpful for story understanding than the most-liked comments in the baseline. Still, one participant (P12) mentioned that the panel-anchored comments can interrupt the reading flow. He commented *"I still prefer to access comments after I finish an episode. Switching narratives is just not for me."*

Room for improvement was noted. Some participants expressed that the interactive features were burdensome if they wanted to use them often, and suggested supporting selection of the level of detail when starting the episode. Also, one participant (P11) pointed out that the presentation order is sometimes unnatural when listening to the on-demand details. She noted *"First I listened to what the character is doing and the conversation, and when I double-tapped, I listened to what he's wearing. The order doesn't make sense sometimes.*

The appropriateness of amount of description (baseline: 5.50 & Cocomix: 5.83) was comparable, showing that participants were satisfied with descriptions generated by following our guidelines. One participant (P17) commented how concise the descriptions were compared to conventional audio descriptions. *"I never listen to audio descriptions for movies because they're just so obsessed with describing everything shown in the scene. But, here I think both versions did the task well."*

7 DISCUSSION

In this section, we reflect on our study findings on adaptive description and interactive description, then discuss how to scale up and generalize our approach. Next, we discuss how broader interaction

Statement ID	Statement	Baseline	Cocomix
S1	I could concentrate during reading.	6.75 (0.45)	6.50 (1.00)
S2	I could process all the given information.	6.42 (0.90)	6.83 (0.39)
S3	The amount of detail was appropriate in each panel	5.50 (1.57)	5.83(1.03)
S4	I had full control over the reading time.	4.83 (1.40)	5.58 (1.38)
S5	**I had full control over the amount of description.***	2.92 (2.23)	**6.50** (0.80)
S6	I am confident in comprehending and feeling the story.	5.08 (1.98)	5.67 (1.37)
S7	I can read stories in efficient manners.	5.67 (1.37)	5.92 (1.38)
S8	**Comments were helpful for understanding the content.***	4.25 (1.76)	**5.25** (1.66)
S9	Comments were fun to read.	6.17 (1.59)	6.42 (1.25)

Table 7: We collected participants' subjective ratings on Cocomix in a 7-point Likert scale. The value represents the mean of scores (SD). Wilcoxon test was run and a statistical significant difference ($p < 0.05$) is marked with *.
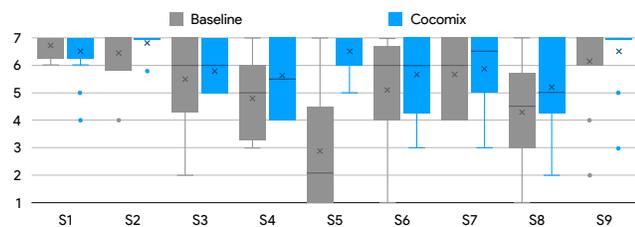
Figure 7: Boxplot of mean scores for the baseline and Cocomix (1 = low, 7 = high). Wilcoxon test was run and a statistical significant difference ($p < 0.05$) is marked with *. Cocomix significantly outperformed the baseline in S5 (Control over the amount of description) and S8 (Usefulness of comments in helping content understanding).

data can power data-drive accessibility. We also discuss limitations of our work, which represent opportunities for future research.

7.1 Adapting Descriptions into Broader Contexts

Cocomix adapts each panel's length of description by how frequently it was mentioned in comments. In the evaluation study, the confidence in story comprehension was comparable in both conditions, implying that our adaptation did not result in any major information loss. However, there was no significant difference in reading efficiency as participants had to adapt the description to their preference in each panel.

There are two dimensions in which we can consider the adaptation of description: 1) adapting the total length of description to user context 2) adapting each panel's length to story context. In this paper, we explored the second dimension. As a future work, we plan to investigate how the description can be adapted under various user contexts. From the usage pattern of detail requests, we found that the desired length depends on individual's preference, and familiarity with the story. From the interview feedback, we also discovered that various reading contexts, such as whether they are in a hurry or relaxing, should be considered so that the description length is adjusted to desired reading time.

To provide appropriate descriptions under various contexts, the mode selection can be provided as suggested by one participant.

As a step forward from 'one-description-fits-all', we envision personalized adaptation of description by leveraging users' interaction patterns.

7.2 Interacting with Description

With Cocomix, BLV users can interact with the description by actively requesting further details and exploring relevant comments. Interestingly, one participant noted that the details would stay longer in memory when it was provided on-demand than granted. This suggests additional applicability of interactive descriptions for highlighting information in addition to providing options.

We distilled considerations for introducing interactivity in descriptions. First, the presentation order should be carefully set, so that the default description alone is sufficient for following the storyline and selective details do not harm the reading flow. This means that not only the dependencies between sentences of a single image (panel), but also dependencies of a sentence on later panels' descriptions should be considered. Prior research [48] has suggested determining the order as how a sighted person sees an image.

Also, the notice of availability for interaction should be carefully designed. During our iterative design process, we discovered that users unaware of availability are concerned about 'unknown unknowns' or frustrated when the requested detail is unavailable. Cocomix provides clear audio cues to inform users that they can explore details or read comments in the panel. However, one unexpected finding was that the notification made it difficult to ignore details for some users, leaving a question of whether BLV users truly had 'freedom' in control.

Finally, the unit of interaction impacts the reading flow. As shown in our results, the concentration measure was slightly lower when BLV readers used the interactive features. We also observed that the desired level of interactivity varied by individuals. For example, one participant suggested sentence-level control to gain complete control over what he receives. Another participant noted the burden of having to make decisions in every panel whether to access details. Interactions in micro-level can support full-control, but its relation to enhanced reading experience is an open question.

7.3 Scalability of Webtoon Descriptions

We proposed guidelines (Section 3.4) for describing webtoons. Yet, there are several challenges with respect to scalability. First, most

webtoons are episodic and periodical thus require consistent efforts for making descriptions. Second, the describer should be familiar with the story context. Prior work has suggested the potential of involving fans to crowdsource comic descriptions [63]. An example of dedicated webtoon fans' contribution is *Naver Webtoon Translate* [12] where fans translate webtoons to share with a global audience. As a globally large population who are familiar with the story and motivated to share the story with others, we believe that webtoon fans have potential to scale up descriptions.

To move toward a more streamlined and automated process, we suggest using templates [26, 46] for ensuring consistency of lengths, expressions, and contents between describers. We believe that data collected can be used to augment future pipelines, such as automatic generation of human-like descriptions. Also, some initial steps, such as grouping continuous panels into *phasels*, can be automated with the advancement of computer vision techniques. While existing research around comics focuses on detection [52, 53, 62] and extraction [13, 16, 56] of sub-elements, we believe higher level of image analysis such as grouping panels into semantically meaningful units or summarization into key panels can further catalyze meaningful interactions with comics.

7.4 Generalizability of Cocomix and Comments-driven Accessibility

We believe our approach on mining comments to improve accessibility can be used in other types of media (e.g., social media images, YouTube videos, etc). To see how our approach can generalize to other types of webcomics, we sampled webcomics from social media and analyzed whether the commenting behavior is different. From Facebook and Instagram, we searched with the query "comics" and selected top-ten webcomics that are regularly published. Then, two researchers analyzed the comments using same taxonomy used in Section 4. We found that more comments are in the form of 'Communication' or 'Self-expression'. It is partially due to the inherent characteristics of the platforms, and also due to 'Mention' feature which readers actively use to share with friends.

The performance of our approach relies on the number of comments. While our guidelines for descriptions are applicable regardless, the core features' effectiveness might not be highlighted well with newly launched webtoons with less viewership. Also, a certain amount of delay is unavoidable to wait for webtoon readers to comment on each episode. Yet, we believe this is not to a great degree as we observed from the collected comments data that majority of readers commented within a few hours each webtoon was uploaded. To better utilize the comments, we make the following suggestions for designers of social features in webcomics platforms: First, support referencing to a panel, or within-panel elements such as characters or objects that are shown to described often (Table 4). This not only increases the accuracy of our linking results, but also encourages readers to leave more specific comments by reducing the burden to specify the target. Second, prompt readers to comment on various elements and attributes. When most readers talk about batman, system can ask *'What do you think about joker's hair?'*. Third, acknowledge and incentivize their prosocial behaviors. For example, active commenters can be provided early access to latest episodes not yet available to public.

7.5 Availability of Other Interaction Data

Reinholt et al. [61] proposed the concept of *use-driven accessibility*, a concept that an image's accessible description can arise from sighted users' interaction with the image, rather than conscious work to create an image description. In this paper, we focused on one type of such data, user comments, which are available at low cost and contain rich contexts. We believe that Cocomix can further benefit from incorporating other interaction data. For example, scroll or gaze data can be analyzed to identify the panels or elements where readers viewed longer, and laughter [35] can be collected to note interesting scenes. However, there remain unanswered questions with leveraging 'byproduct' data such as "Will the output align with the creator's intent?", and "Can we expect the same focus-patterns from BLV readers as sighted readers?" We hope our research catalyzes future work in data-driven accessibility.

7.6 Limitations and Future Work

Our work has several limitations which we address in this section.

- We did not explore how peculiarities of comics medium such as aesthetics can be conveyed well in the description. From our formative study, we discovered that most participants were not interested in the drawing style or the layout of panels as readers born with visual impairments lacked the sensory experience. To support readers to picture the scene more quickly as they do with drawings, it would be enriching to complement descriptions with audio augmentations.
- We did not investigate the effectiveness of our webtoon description guidelines in terms of unambiguity and output consistency. Our next step is to assess and improve our guidelines with feedback from both professional and non-professional describers.
- Cocomix defines *key panels* as panels with elements frequently mentioned in comments. Thus, it has limited capability of detecting panels with setting introductions which play a significant role in understanding forthcoming panels but less mentioned in comments. One solution may be to consider other data such as time spent on each panel, or dependencies between panel-descriptions to cover various context of 'importance'.
- Both studies involved limited number of participants (formative study: 10, evaluation study: 12). We need to better account for the diverse needs of people with different degrees of sight loss, usages of assistive tools, and technical skills [20, 54]. Our pipeline evaluation only covered five different webtoons and we need to further explore the pipeline's robustness in diverse webtoons. Also, we evaluated extracted comments individually, not rating how diverse information they provide as a whole.
- We did not explore how users perceive and react with incorrect output of the system as no user reported that they noticed irrelevant comments in our user study. While we

---

[12]https://translate.webtoons.com/guide

showed the feasibility of pipeline through separate evaluation, incorrect links can still exist where human validation such as reporting from users would be valued.

8 CONCLUSION

We present Cocomix, an interactive webtoon reader that utilizes comments to improve non-visual webtoon accessibility. Cocomix provides adaptive description through comments-based scoring and abridging, and supports easy access to relevant comments through panel-anchored comments. We hope our research catalyzes future work in making broader types of visual media accessible. We also encourage web platforms to take note of the success of these comments-driven interaction methods and integrate them to improve non-visual accessibility.

ACKNOWLEDGMENTS

REFERENCES

[1] Dragan Ahmetovic, Nahyun Kwon, Uran Oh, Cristian Bernareggi, and Sergio Mascetti. 2021. Touch Screen Exploration of Visual Artwork for Blind People. In *Proceedings of the Web Conference 2021*. 2781–2791.

[2] Kholoud Khalil Aldous, Jisun An, and Bernard J Jansen. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 47–57.

[3] Chad Allen. [n.d.]. *Unseen*. Retrieved August 30, 2021 from https://www.unseencomic.com/

[4] Alphatart&Sumpul. [n.d.]. *The Remarried Empress*. Retrieved September 2, 2021 from https://www.webtoons.com/en/fantasy/the-remarried-empress/list?title_no=2135

[5] Alphatart/Sumpul. [n.d.]. *The Remarried Empress*. Retrieved August 30, 2021 from https://www.webtoons.com/en/fantasy/the-remarried-empress/list?title_no=2135

[6] Kohei Arai and Herman Tolle. 2010. Automatic e-comic content adaptation. *International Journal of Ubiquitous Computing* 1, 1 (2010), 1–11.

[7] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2016. A simple but tough-to-beat baseline for sentence embeddings. (2016).

[8] artweb. [n.d.]. *DESCRIVEDDO GUIDELINES*. Retrieved September 9, 2021 from https://artweb.netlify.app/desc_en

[9] asimplebengo. [n.d.]. *Hybrid*. Retrieved September 2, 2021 from https://www.webtoons.com/en/challenge/hybrid/list?title_no=211861

[10] audiobook. [n.d.]. *a movie in your mind*. Retrieved August 30, 2021 from https://www.graphicaudiointernational.net/

[11] Olivier Augereau, Motoi Iwata, and Koichi Kise. 2018. A survey of comics research in computer science. *Journal of imaging* 4, 7 (2018), 87.

[12] braille comics. [n.d.]. *Laville Braille*. Retrieved August 30, 2021 from http://www.lavillebraille.fr/des-livres-a-voir-et-a-toucher/

[13] João MC Correia and Abel JP Gomes. 2016. Balloon extraction from complex comic books using edge detection and histogram scoring. *Multimedia Tools and Applications* 75, 18 (2016), 11367–11390.

[14] Jakob Dittmar. 2014. Comics for the blind and for the seeing. *International Journal of Comic Art; 1* 16 (2014).

[15] David Dubray and Jochen Laubrock. 2019. Deep CNN-based speech balloon detection and segmentation for comic books. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1237–1243.

[16] Arpita Dutta and Samit Biswas. 2019. CNN based extraction of panels/characters from bengali comic book page images. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Vol. 1. IEEE, 38–43.

[17] Teresa Kardoulias Sarah Stephenson Keyes Elisabeth Salzhauer Axel, Virginia Hooper and Francesca Rosenberg. [n.d.]. *AEB's Guidelines for Verbal Description*. Retrieved September 9, 2021 from http://www.artbeyondsight.org/handbook/acs-guidelines.shtml

[18] EmiMG. [n.d.]. *ZomCom*. Retrieved September 2, 2021 from https://www.webtoons.com/en/challenge/zomcom/list?title_no=70195

[19] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1175–1184.

[20] Anita Fidyka and Anna Matamala. 2018. Audio description in 360° videos. *Translation Spaces* (2018).

[21] Pet Foolery. [n.d.]. *Pixie and Brutus*. Retrieved September 2, 2021 from https://www.webtoons.com/en/challenge/pixie-and-brutus/list?title_no=452175

[22] Pet Foolery. [n.d.]. *Pixie and Brutus*. Retrieved August 30, 2021 from https://www.webtoons.com/en/challenge/pixie-and-brutus/list?title_no=452175

[23] Benjamin Fraser. 2020. Tactile comics, disability studies and the mind's eye: on "A Boat Tour"(2017) in Venice with Max. *Journal of Graphic Novels and Comics* (2020), 1–13.

[24] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[25] Cole Gleason, Amy Pavel, Himalini Gururaj, Kris Kitani, and Jeffrey P Bigham. 2020. Making GIFs Accessible.. In *ASSETS*. 24–1.

[26] Cole Gleason, Amy Pavel, Xingyu Liu, Patrick Carrington, Lydia B Chilton, and Jeffrey P Bigham. 2019. Making memes accessible. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 367–376.

[27] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M Kitani, and Jeffrey P Bigham. 2020. Twitter A11y: A browser extension to make Twitter images accessible. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[28] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 1373–1378. https://aclweb.org/anthology/D/D15/D15-1162

[29] Mohit Iyyer, Varun Manjunatha, Anupam Guha, Yogarshi Vyas, Jordan Boyd-Graber, Hal Daume, and Larry S Davis. 2017. The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*. 7186–7195.

[30] Jakka. [n.d.]. *Independant Diary*. Retrieved August 30, 2021 from https://comic.naver.com/webtoon/list?titleId=748105&no=78&weekday=thu

[31] JH. [n.d.]. *The Boxer*. Retrieved August 30, 2021 from https://comic.naver.com/webtoon/list?titleId=736989&weekday=thu

[32] Violet Karim. [n.d.]. *Familiar Feelings*. Retrieved September 2, 2021 from https://www.webtoons.com/en/challenge/familiar-feelings/list?title_no=323558

[33] Hyunwoo Kim, Haesoo Kim, Kyung Je Jo, and Juho Kim. 2021. StarryThoughts: Facilitating Diverse Opinion Exploration on Social Issues. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

[34] Jeffrey SJ Kirchoff. 2013. It's just not the same as print (and it shouldn't be): Rethinking the possibilities of digital comics. *Technoculture: An Online Journal of Technology in Society* 3, 1 (2013).

[35] Soyoung Kwon and Kun-Pyo Lee. 2016. What makes readers laugh? value of sensing laughter for humor webtoon. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 867–874.

[36] Yunjung Lee, Hwayeon Joh, Suhyeon Yoo, and Uran Oh. 2021. AccessComics: an accessible digital comic book reader for people with visual impairments. In *Proceedings of the 18th International Web for All Conference*. 1–11.

[37] Luyuan Li, Yongtao Wang, Liangcai Gao, Zhi Tang, and Ching Y Suen. 2014. Comic2CEBX: A system for automatic comic content adaptation. In *IEEE/ACM Joint Conference on Digital Libraries*. IEEE, 299–308.

[38] Guanhong Liu, Xianghua Ding, Chun Yu, Lan Gao, Xingyu Chi, and Yuanchun Shi. 2019. " I Bought This for Me to Look More Ordinary" A Study of Blind People Doing Online Shopping. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[39] Yang Liu. 2019. Fine-tune BERT for extractive summarization. *arXiv preprint arXiv:1903.10318* (2019).

[40] Zongyang Ma, Aixin Sun, Quan Yuan, and Gao Cong. 2012. Topic-driven reader comments summarization. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 265–274.

[41] Scott McCloud. 1993. Understanding comics: The invisible art. *Northampton, Mass* (1993).

[42] Chris McCoy. [n.d.]. *Safely*. Retrieved September 2, 2021 from https://www.webtoons.com/en/comedy/safely-endangered/list?title_no=352

[43] Mimino666. [n.d.]. *langdetect: Port of Google;s language-detection library*. Retrieved September 9, 2021 from https://github.com/Mimino666/langdetect#languages

[44] Elaheh Momeni, Claire Cardie, and Myle Ott. 2013. Properties, prediction, and prevalence of useful user-generated comments for descriptive annotation of

social media objects. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7.

[45] Elaheh Momeni, Ke Tao, Bernhard Haslhofer, and Geert-Jan Houben. 2013. Identification of useful user comments in social media: A case study on Flickr commons. In *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*. 1–10.

[46] Valerie S Morash, Yue-Ting Siu, Joshua A Miele, Lucia Hasty, and Steven Landau. 2015. Guiding novice web workers in making image descriptions using templates. *ACM Transactions on Accessible Computing (TACCESS)* 7, 4 (2015), 1–21.

[47] Giulio Mori, Maria Claudia Buzzi, Marina Buzzi, and Barbara Leporini. 2010. Structured audio podcasts via web text-to-speech system. In *Proceedings of the 19th international conference on World wide web*. 1281–1284.

[48] Meredith Ringel Morris, Jazette Johnson, Cynthia L Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.

[49] Murrz. [n.d.]. *Murrz*. Retrieved September 2, 2021 from https://www.webtoons.com/en/slice-of-life/murrz/list?title_no=1281

[50] Feng Nan, Ramesh Nallapati, Zhiguo Wang, Cicero Nogueira dos Santos, Henghui Zhu, Dejiao Zhang, Kathleen McKeown, and Bing Xiang. 2021. Entity-level Factual Consistency of Abstractive Text Summarization. *arXiv preprint arXiv:2102.09130* (2021).

[51] Anime News Network. [n.d.]. *Japanese Volunteers Transcribe Manga for Blind People*. Retrieved August 30, 2021 from https://www.animenewsnetwork.com/news/2007-07-24/japanese-volunteers-transcribe-manga-for-blind-people

[52] Nhu-Van Nguyen, Christophe Rigaud, and Jean-Christophe Burie. 2017. Comic characters detection using deep learning. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, Vol. 3. IEEE, 41–46.

[53] Toru Ogawa, Atsushi Otsubo, Rei Narita, Yusuke Matsui, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2018. Object detection for comics using manga109 annotations. *arXiv preprint arXiv:1803.08670* (2018).

[54] Pilar Orero, Stephen Doherty, Jan-Louis Kruger, Anna Matamala, Jan Pedersen, Elisa Perego, Pablo Romero-Fresco, Sara Rovira-Esteva, Olga Soler-Vilageliu, and Agnieszka Szarkowska. 2018. Conducting experimental research in audio-visual translation (AVT): A position paper. *JosTrans: The Journal of Specialised Translation* 30 (2018), 105–126.

[55] Rachel Sarah Osolen and Leah Brochu. 2020. Creating an Authentic Experience. *The International Journal of Information, Diversity, & Inclusion (IJIDI)* 4, 1 (2020).

[56] Xiaoran Qin, Yafeng Zhou, Yonggang Li, Siwei Wang, Yongtao Wang, and Zhi Tang. 2019. Progressive deep feature learning for manga character recognition via unlabeled training data. In *Proceedings of the ACM Turing Celebration Conference-China*. 1–6.

[57] Vipul Raheja and Joel Tetreault. 2019. Dialogue act classification with context-aware self-attention. *arXiv preprint arXiv:1904.02594* (2019).

[58] Frédéric Rayar. 2017. Accessible comics for visually impaired people: Challenges and opportunities. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 3. IEEE, 9–14.

[59] Frédéric Rayar, Bernard Oriola, and Christophe Jouffrais. 2020. ALCOVE: an accessible comic reader for people with low vision. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 410–418.

[60] Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi, and Meredith Ringel Morris. 2017. Eyes-free art: Exploring proxemic audio interfaces for blind and low vision art engagement. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[61] Kyle Reinholt, Darren Guinness, and Shaun K Kane. 2019. Eyedescribe: Combining eye gaze and speech to automatically create accessible touch screen artwork. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. 101–112.

[62] Christophe Rigaud. 2014. *Segmentation and indexation of complex objects in comic book images*. Ph.D. Dissertation. Université de La Rochelle.

[63] Christine Samson, Casey Fiesler, and Shaun K Kane. 2016. "Holy Starches Batman!! We are Getting Walloped!" Crowdsourcing Comic Book Transcriptions. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 289–290.

[64] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. 2010. How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*. 891–900.

[65] Rachel Smythe. [n.d.]. *Lore Olympus*. Retrieved September 2, 2021 from https://www.webtoons.com/en/romance/lore-olympus/list?title_no=1320

[66] Rachel Smythe. [n.d.]. *Lore Olympus*. Retrieved August 30, 2021 from https://www.webtoons.com/en/romance/lore-olympus/list?title_no=1320

[67] Abigale J Stangl, Esha Kothari, Suyog D Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. Browsewithme: An online clothes shopping assistant for people with visual impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 107–118.

[68] Raymond Hendy Susanto, Hai Leong Chieu, and Wei Lu. 2016. Learning to capitalize with character-level recurrent neural networks: an empirical study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2090–2095.

[69] tactile comics. [n.d.]. *Life - A tactile comic for blind people | Philipp Meyer*. Retrieved August 30, 2021 from https://www.hallo.pm/life/

[70] Carla Tamburro, Timothy Neate, Abi Roper, and Stephanie Wilson. 2020. Accessible Creativity with a Comic Spin. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–11.

[71] Garreth W Tigwell, Benjamin M Gorman, and Rachel Menzies. 2020. Emoji Accessibility for Visually Impaired People. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[72] Web Accessibility Tutorials. [n.d.]. *Images Concepts - Images - WAI*. Retrieved September 9, 2021 from https://www.graphicaudiointernational.net/

[73] uru chan. [n.d.]. *unOrdinary*. Retrieved September 2, 2021 from https://www.webtoons.com/en/super-hero/unordinary/episode-223/viewer?title_no=679&episode_no=234

[74] Violeta Voykinska, Shiri Azenkot, Shaomei Wu, and Gilly Leshed. 2016. How blind people interact with visual content on social networking services. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 1584–1595.

[75] Ruolin Wang, Zixuan Chen, Mingrui Ray Zhang, Zhaoheng Li, Zhixiu Liu, Zihan Dang, Chun Yu, and Xiang'Anthony' Chen. 2021. Revamp: Enhancing Accessible Information Seeking Experience of Online Shopping for Blind or Low Vision Users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[76] Xinwei Wang, Jun Hu, Bart Hengeveld, and Matthias Rauterberg. 2018. Segmentation of panels in d-Comics. In *Interactivity, Game Creation, Design, Learning, and Innovation*. Springer, 28–37.

[77] Xinwei Wang, Jun Hu, Bart Hengeveld, and Matthias Rauterberg. 2019. Expressing segmentation in d-comics. In *International Conference on Human-Computer Interaction*. Springer, 402–409.

[78] Shaomei Wu and Lada A Adamic. 2014. Visually impaired users on an online social network. In *Proceedings of the sigchi conference on human factors in computing systems*. 3133–3142.

[79] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1180–1192.

[80] Yaongyi. [n.d.]. *True Beauty*. Retrieved September 2, 2021 from https://www.webtoons.com/en/romance/truebeauty/list?title_no=1436

[81] Matin Yarmand, Dongwook Yoon, Samuel Dodson, Ido Roll, and Sidney S Fels. 2019. "Can you believe [1: 21]?!" Content and Time-Based Reference Patterns in Video Comments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[82] Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Peng Li, Maosong Sun, and Zhiyuan Liu. 2020. Coreferential reasoning learning for language representation. *arXiv preprint arXiv:2004.06870* (2020).

[83] Jeffrey M Zacks, Barbara Tversky, and Gowri Iyer. 2001. Perceiving, remembering, and communicating structure in events. *Journal of experimental psychology: General* 130, 1 (2001), 29.

A STUDY PARTICIPANTS DEMOGRAPHICS

PID	Age	Gender	Visually impairment description	Onset	Experienced comics before or after vision loss
P1	22	Female	Totally blind	Congenital	Never
P2	22	Male	Low vision	Acquired	Before&After
P3	22	Female	Totally blind	Congenital	Never
P4	23	Male	Low vision	Acquired	Before&After
P5	26	Male	Totally blind	Acquired	Before
P6	28	Male	Totally blind	Congenital	Never
P7	29	Female	Totally blind	Congenital	Never
P8	35	Female	Totally blind	Congenital	Never
P9	25	Male	Totally blind	Acquired	Never
P10	28	Male	Low vision	Congenital	Never
P11	23	Male	Totally blind	Congenital	After
P12	53	Male	Totally blind	Acquired	Before
P13	29	Male	Totally blind	Congenital	Never
P14	23	Female	Low vision	Acquired	Never
P15	28	Female	Totally blind	Congenital	Before&After
P16	35	Male	Totally blind	Congenital	After
P17	23	Female	Low vision	Acquired	Never
P18	22	Female	Totally blind	Congenital	Never
P19	29	Male	Totally blind	Congenital	Never
P20	36	Male	Totally blind	Congenital	Before&After
P21	36	Male	Totally blind	Congenital	Never
P22	27	Male	Low vision	Acquired	Never

Table 8: Study Participants (P1-P10: Formative study participants, P11-P22: Evaluation study participants. All participants are screen reader users.)