Making Memes Accessible

Cole Gleason
Carnegie Mellon University
Pittsburgh, USA
cgleason@cs.cmu.edu

Amy Pavel
Carnegie Mellon University
Pittsburgh, USA
apavel@cs.cmu.edu

Xingyu Liu
Carnegie Mellon University
Pittsburgh, USA
xingyul@andrew.cmu.edu

Patrick Carrington
Carnegie Mellon University
Pittsburgh, USA
pcarrington@cmu.edu

Lydia B. Chilton
Columbia University
New York, USA
chilton@cs.columbia.edu

Jeffrey P. Bigham
Carnegie Mellon University
Pittsburgh, USA
jbigham@cs.cmu.edu

ABSTRACT
Images on social media platforms are inaccessible to people with vision impairments due to a lack of descriptions that can be read by screen readers. Providing accurate alternative text for all visual content on social media is not yet feasible, but certain subsets of images, such as internet memes, offer affordances for automatic or semi-automatic generation of alternative text. We present two methods for making memes accessible semi-automatically through (1) the generation of rich alternative text descriptions and (2) the creation of audio macro memes. Meme authors create alternative text templates or audio meme templates, and insert placeholders instead of the meme text. When a meme with the same image is encountered again, it is automatically recognized from a database of meme templates. Text is then extracted and either inserted into the alternative text template or rendered in the audio template using text-to-speech. In our evaluation of meme formats with 10 Twitter users with vision impairments, we found that most users preferred alternative text memes because the description of the visual content conveys the emotional tone of the character. As the preexisting templates can be automatically matched to memes using the same visual image, this combined approach can make a large subset of images on the web accessible, while preserving the emotion and tone inherent in the image memes.

Author Keywords
alternative text, meme, blind, low vision, audio, social media, image description

INTRODUCTION
Increasingly, people communicate on social media networks and in personal chats using visual content (*e.g.*, emojis, memes, and recorded images/videos). However, a large amount of the visual content on social media networks and personal chats remains inaccessible due to a lack of high-quality image descriptions. Social media platforms like Facebook [35], Twitter [30], and Instagram [15] allow users to add alternative text to their images, but most do not use this feature resulting in only 0.1% of images becoming accessible [10]. Because social media platforms and users do not include high-quality alt text with all images, we explore how to exploit repetition in the common content users share over time. A large number of images shared on social media are not original images. In fact, a recent study of images on Twitter revealed that of a sample of over 1.7 million photos, 80% were retweeted images [10]. In this paper, we focus on a class of image content which affords opportunities to leverage this repetition – memes.

Broadly, a meme is "an idea, behavior, or style that spreads from person to person within a culture – often with the aim of conveying a particular phenomenon, theme, or meaning represented by the meme"[1]. We focus on image macro memes [8], a common form of image-based meme that features an image overlaid with caption text (Figure 2). Sharing an iden-

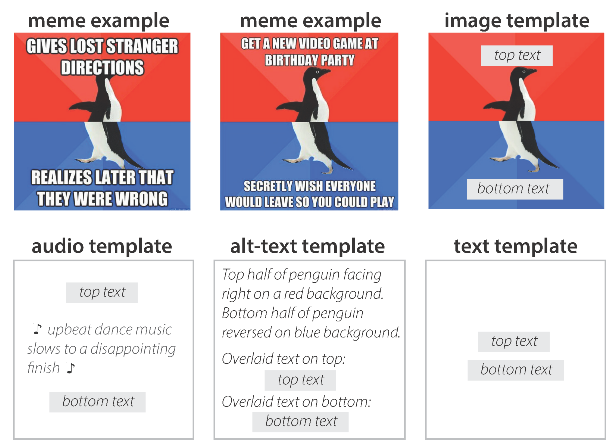

Figure 1. Image macro memes feature a meme example that can be described with an image template. We propose alternative forms of meme description including audio, alt-text, and text templates.

[1]https://www.merriam-webster.com/dictionary/meme

ASSETS '19 October 28–30, 2019, Pittsburgh, PA, USA

ISBN 978-1-4503-6676-2/19/10. . . $15.00
DOI: http://dx.doi.org/10.1145/3308561.3353792

Figure 2. Examples of image macro memes from two image templates. Template A represents the "Success Kid" meme and Template B represents the "First World Problems" meme.

tifiable image macro meme can serve as shorthand for "a phenomenon, theme or meaning". For example, the celebrating toddler image represents "common situations with minor victories" (Figure 2A), and the crying woman image represents "first world problems" (Figure 2B). However, existing alt text for image macro memes typically describe only the meme text (*e.g.*, "Put candy bar in shopping cart without mom noticing"), dropping the relevant context provided by the template. Without the context recognized through the images, the memes often lose their emotional tone or humorous aspect.

To make memes more widely accessible, we propose 1) an automatic method for applying existing image descriptions to new meme examples, and 2) a non-expert workflow for creating high-quality alt text and audio macro meme templates. Our automatic workflow classifies a meme example with 92% accuracy and recognizes meme example text with a 22% word error rate (9.2% by character error rate).

To understand user preferences for an accessible meme format, we conducted a user study with 10 visually impaired participants comparing 3 different meme formats: meme text only, image description with meme text, and meme text with a unique tonally-relevant background sound (created by a sound designer). While users preferred image descriptions, we find that our traditional image descriptions occasionally fail to efficiently convey the function of the image (*e.g.*, shorthand for tone). For audio, despite quickly conveying tone, a background sound can lack universal accessibility. Based on user performance and preference, we propose structured questions for creating image descriptions for image macro memes.

In summary, our contributions are as follows:

- An automatic process to recognize known memes and extract new text,
- An interface for creating accessible memes in alternative text or audio formats with placeholders for the extracted text, and
- Structured questions to be used for alternative description formats for visual image content, specifically memes.

RELATED WORK

Our investigation of the accessibility of memes is related to prior research on the usage of alternative text, methods to generate image descriptions, and characterization and use of memes on the web.

Online Accessibility for Images

Alternative text or "alt text", most commonly refers to captions for images online or in other software. The text is typically added by website developers when creating web pages, either in the HTML source code or via web content creation tools. Today, accessibility for screen-reader users is one of the most commonly cited reasons to add alternative text to images, however, it is also useful for non-graphical browsers or when an image does not load for sighted users [1, 5]. In fact, image descriptions have been used for a number of different applications including "semantic visual search, visual intelligence in chatting robots, photo and video sharing in social media, and aid for visually impaired people to perceive surrounding visual content" [13]. Image labels, captions, and descriptions provide a solid foundation for many of these kinds of applications. In this paper, we focus on the communicative qualities of visual content alternatives for human users [26]. For the most part, on the web this means either text or audio descriptions for images and videos. Alt text descriptions for images have been standard since 1995, but recent research by Morris *et al.* contends that this standard may be stale, and modern computing platforms could support richer representation of visual content, including audio [22].

A historical analysis of websites reveals that complexity and accessibility have had an inverse relationship; as websites become more complex, they have become less accessible [12]. Guinness *et al.* created a system to identify and provide missing alternative text based on similar images found on the web. After using Caption Crawler, 20-35% of images on various categories of websites were still lacking alt text [11]. With the rise of social media platforms, a significant amount of image content on the web is now generated by end-users, not website authors. This has led to a large amount of content being inaccessible, as users either did not have the option to add descriptions to their posts or were ill-equipped to produce high quality descriptions [10].

Alt Text Generation

The majority of alternative text is written manually by website developers or authors of the website content. While authors are recommended to follow Web Content Accessibility Guidelines [4], many images on the web are not labelled correctly or at all [2, 12]. Researchers have since sought to automatically generate descriptions for images on the web. Different approaches have been adopted in order to label both objects [35] and descriptions of scenes [20]. However, both the labelling techniques and descriptions should be accepted cautiously, as prior work has also highlighted a quality threshold for where the generated descriptions can do more harm than good [26].

The best alternative text is typically provided by human labelling, especially for complex photos with a specific intent. Researchers have proposed methods of sharing alternative text for images between users [6] and collaboratively make websties more accessible without permission from the owner [28]. Guinness *et al.* proposed the Caption Crawler to automatically retrieve alternative text attached to the same image elsewhere on the web through reverse image searches, which achieves a similar goal without active crowdsourcing [11]. As this re-uses alt text from the same image across the web, it is a poor approach for memes which are visually similar but with different meanings depending on the overlaid text.

Memes and Humor

Meme are challenging to describe in alternative text because they contain humor. According to the Semantic Script Theory of Humor [24], what is communicated in humor is *implied* rather than stated directly. According to this theory, jokes have a set-up and a punchline: the set-up leads the listener to expect one thing, but then the punchline violates that expectation and forces the listener to think of a second interpretation that connects both statements. Often the second interpretation involves an insult or an error in logic [19]. For example, in Figure 2, the "Success Kid" meme (Template A) has set-up text at the top saying "[I] put candy in the shopping cart", which is a normal thing to do. Then there is a picture of a toddler looking very proud of himself, and a punchline reading "without [my] mom noticing." This implies he did it sneakily and he is proud that his mischievous act was not punished. Additionally, the speaker is exaggerating how big this accomplishment is. It is relatively minor, but the serious look of success on the kids face implies he is treating it as a big accomplishment. This is the error in logic, and perhaps a self-effacing insult that is meant to make it humorous to the reader.

Understanding humor relies on a shared context of the speaker and the listener in order for the listener to infer the correct meaning. This is difficult for both people and computers. Although many computer programs have been trained to detect humor, most struggle to achieve more than 80% accuracy over a 50% baseline [27, 29, 16, 3, 21]. This is likely because of the immense amount of cultural background as well as necessary ability to interpret the hidden meaning that is required. Additionally, people outside of a culture context often find that culture's humor difficult to understand. A study of people unfamiliar with memes or meme subculture [18] found that memes were very hard to understand. They tested several ways of elaborating or explaining the memes and found the most successful strategy was to provide crowdsourced annotations which explicitly described the implied meaning according to the Semantic Script Theory of Humor. As noted by the common quotation [34], "Humor can be dissected, as a frog can, but the thing dies in the process and the innards are discouraging to any but the pure scientific mind." In this vein, there is a challenge in making the content of a meme more accessible, while still leaving the meaning implied, so that the joke can be enjoyed as intended.

MAKING MEMES ACCESSIBLE

To transform image macro memes into accessible alternative formats, we provide 1) an *automatic method* for converting image macro memes encountered on the web into alternative meme formats, 2) an *authoring interface* for generating meme alt text templates and audio macro meme templates. As each meme template can apply to thousands of instances of the same base meme, our automatic method allows people browsing the web to convert existing image macro memes to preexisting alternative meme template formats (*e.g.*, meme text, alt text, audio meme). Our authoring interface enables non-experts to efficiently produce meme template alternatives.

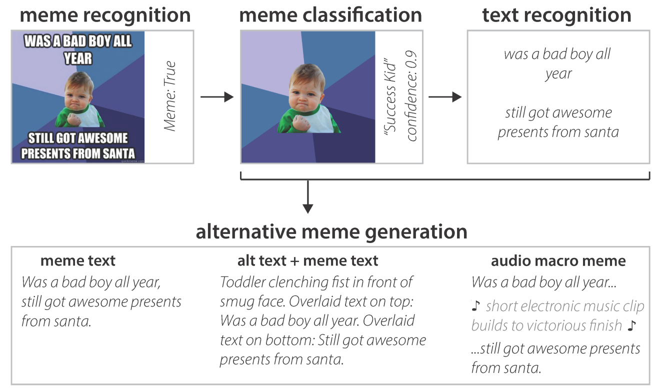

Figure 3. Our system first recognizes whether or not the image is a meme. If it is a meme, the system attempts to classify the meme as a representative example of a meme template in our database (*e.g.*, "Success Kid"), and recognizes the text within the meme (*e.g.* "Was a bad boy all year"). If the meme classification confidence for a match (i.e. image similarity score) reaches a score over a given threshold, we output three formats: meme text only, an alt text + meme text pair, and an audio macro meme. If the confidence falls below that threshold, we output only the text.

Automatic method

We automatically convert existing image macro memes encountered in the wild to alternative meme types by: 1) recognizing that an image is a meme, 2) identifying the meme type (*e.g.*, "success kid", "confession bear"), and 3) extracting the text from the meme (Figure 3). We then insert the extracted text into the alternative text templates textually or audio macro meme template using text to speech.

Meme recognition

When a user encounters an image on a social media network (*e.g.*, Imgur, Twitter), we first detect whether or not the image is a meme using Google Cloud Vision API's "Detecting Web Entities and Pages" request. For a given image, we obtain a list of web-generated labels (*e.g.* "Meme, Success Kid, Toddler, Brother" for the Success Kid meme) and we check if the keyword "meme" or "internet meme" appears in the list of labels. We evaluated this method with 105 meme images randomly selected from the "Meme Generator Dataset" from Library of Congress's Web Archive [23], and 105 non-meme images (a random subset of the ImageNet database [7]). This method achieves a meme recognition accuracy of 94.4% (100% precision, 89.9% recall). The API typically does not include the "meme" label for new or less prevalent memes.

Meme classification

We next match the recognized input meme to a meme template in order to identify any corresponding alternative meme representation. We create a dataset of the 137 meme templates from Imgur[2]. To automatically match the input meme image with a database meme template, we first re-size and crop the input meme image to be the same size as the templates in the database. Then, we compute for the input meme and each database meme template: 1) the structural similarity between the input image and the template image, and 2) the color histogram difference between the input image and the template image. To compute structural similarity, we use the Multi-Scale Structural Similarity (MS-SSIM) index [33] that considers the luminance, contrast, and structural similarity between image regions at various zoom levels. To compute the color histogram difference, we divide each image into 5 regions (Figure 4) and sum together chi-squared distance between HSV color histograms computed for each region (8 bins for the hue channel, 12 bins for the saturation channel and 3 bins for the value channel) [25]. We define the final image similarity score between two images X and Y as: $\alpha MSSSIM(X,Y) - \beta COLORDIFF(X,Y)$, where α and β are adjustable parameters that sum to 1. We use $\alpha = 0.15$ and $\beta = 0.85$, determined empirically. We calculate an Image Similarity score for each template with the fixed input meme example, and return the template with the highest similarity score. If the score is below a confidence threshold, we only output the meme text, as it is likely not in the database.

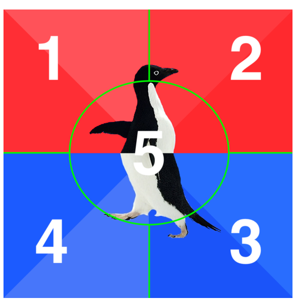

Figure 4. An example of separate regions computed for the color histogram difference measurement.

We evaluated meme classification with 385 memes scraped from the "most popular memes of the year" page of Imgur[3]. With the structural similarity (MS-SSIM) score alone, we achieve an accuracy of 79.22%. The structural similarity score method tends to not perform well on images with low resolution or noise, and performs well on photographs with high-contrast. The color histogram difference alone achieves an accuracy of 77.58%. The color histogram difference method often confuses images with similar colors in the same regions (*e.g.*, the nose of a black bear with a black t-shirt). The combined Image Similarity accuracy is 92.25%.

Text Recognition

After we match the input meme image to a meme template, we extract the top and bottom caption text of the meme image (Figure 2). Given the extracted text and recognized meme template, we can 1) generate the meme's alternative text, and 2) generate an audio meme by using text to speech. We use Google Cloud Vision API's Optical Character Recognition (OCR) feature to detect and extract text from images. Most of the watermarks on memes (*e.g.*, "Imgur.com") appear along image boundaries but do not contribute to the main meme text. So, we remove any text with a bounding box within 5 pixels of the image border.

We evaluated our this recognition approach using the "Meme Generator Dataset" from Library of Congress's Web Archive [23] that contains 57,000 memes along with the top and bottom text. For each ground truth and prediction pair, we calculate word error rate (WER) or the number of substitutions, deletions and insertions in an edit distance alignment over the total number of words [32]. We achieve a word error rate of 22.1% and a character error rate of 9.2%. We find two common types of errors: 1) a word includes only a few mistaken characters ("OET" instead of "GET"), and 2) two words are recognized as one word ("ANDTWO" instead of "AND TWO"). When a word is not recognized, a screen reader either pronounces the word phonetically or spells out the word. In the case of combined words, the phonetic pronunciation is typically correct. We explored applying a simple spell-checker to the resulting OCR text. While it did correct many 1-character mistakes, it often incorrectly changed the combined words. We chose not to use the spell-checker, but in future work we will explore more approaches to reduce the WER, such as spell checkers with more advanced language models or OCR fine-tuned for fonts typically used in image macro memes.

Authoring Alternative Meme Templates

Our authoring interface (Figure 5) lets users generate alternative templates including alt text templates and audio meme templates to add to the database.

The authoring interface accepts an input example meme (Figure 5A) and parses the meme using the automatic pipeline to identify the top or bottom text. To create an alt text template, a user drags the (Figure 5D) top/bottom text placeholders to the meme template box and writes alt text in relation to where it should occur to the placeholders. The system then exports the template as text such that the automatic method can later apply the template to new examples. To create an audio macro meme, a user can place top/bottom text placeholders then click and drag (Figure 5E) sounds from a library accessed via search to place sounds in relation to the placeholders. Finally users can optionally place (Figure 5F) pauses for comedic timing.

The authoring interface is the same for creating either alt text or audio meme templates, except that sounds and pauses are unavailable for alt text meme templates. Authoring of the meme template occurs for the general instance of that meme, so users cannot edit OCR results that will eventually fill the placeholder. However, they can preview their alt text or audio templates with an example.

Once a user has created and submitted their new alt text or audio template, it is reused for any user after a meme example is matched to that base meme template. The system currently chooses just the most recent template, but future work may involve a measure of popularity or voting to assign a default alt text or audio template to a meme.

[2]https://imgur.com/memegen/
[3]https://imgur.com/memegen/popular/year

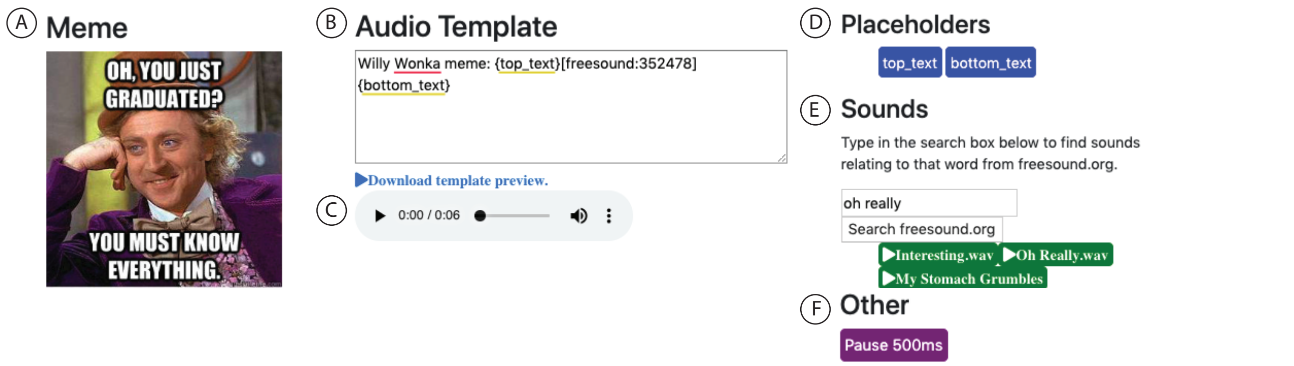

Figure 5. The meme template creation interface displays (A) a reference meme example, (B) the constructed meme template so far, (C) preview and output in text and audio formats, and then a series of tools to construct the meme template. To create an alt text template, a user can drag the (D) top/bottom text placeholders to the meme template box then write alt text in relation to where it should occur with the placeholders. The system then exports the template as text to be applied by the automatic method. To create an audio macro meme, a user can input placeholders then click and drag (E) sounds from a library accessed via search to place sounds in relation to the placeholders. Finally, users can optionally place (F) pauses for comedic timing.

The authoring interface itself is not currently accessible to screen readers, as it is designed to translate visual content, and also relies heavily on drag-and-drop interactions. In future work, we intend to explore accessible interfaces for designing audio-first or alt text-first memes, in addition to translating image macro memes.

MEME FORMAT EVALUATION

We conducted a user study and interview with 10 blind or low-vision participants to understand their experiences with internet memes and compare different media formats to make them accessible. Eleven participants were recruited on the Twitter platform, and participated in our study remotely over online voice chat or phone. One participant (P8) was unable to complete the study due to issues with audio on their computer, so their data is excluded from these results. Participant ages ranged from 19 to 53, with an average age of 31.8. Three participants were female and seven were male. All participants accessed Twitter using a screen reader. All participants reported they had encountered memes before. But, due to accessibility issues with memes, only two participants reported experiencing memes in more depth: P6 reported friends explaining memes, and P9 experienced accessible memes on sites like Instagram. Further participant demographics can be found in Table 1.

Meme Formats

The participants in our study were asked to interact with meme examples sourced from Imgur and Meme Generator's list of popular memes [9]. There were 9 different meme types (Appendix A), with 5 examples of each, for a total of 45 meme examples. The participants experienced 15 examples of these memes in the following three conditions:

1. **Text Only:** As a baseline, the simplest media format was the text-only results from an automatic OCR pass of the meme. These were HTML images that contained alternative text of only the overlaid text. If memes have any alt text at all, it is common for it to only be the overlaid text that the meme generator automatically added. This also represents a completely automatic solution without human involvement, but the visual elements from the image are lost in these descriptions.
2. **Meme Description:** The alternative text in this condition contained a description of the visual content of the image and the overlaid text. The text was separated by the top an bottom of the image, so the participant could tell how they were visually separated.
3. **Audio Macro Memes:** Visual memes intend to provoke an emotional reaction, often some form of humor, that is lost in a pure textual description read by a screen reader. Audio macro memes, a sound analog to image macro memes, include background sound that can carry the emotional affect the meme creator intended. These were sound files that contained background audio customized to each meme type. Text-to-speech rendered the overlaid text in the meme. We hired a professional sound producer to create these audio versions, attempting to convey the emotional tone of the visual meme.

The examples we presented (Appendix A) represented a best case scenario in quality of meme examples. For all of these memes, we corrected the OCR results before generating each example, in order to ensure participants were evaluating the meme formats, not the OCR results. Members of the research team who were familiar with alternative text wrote the image descriptions for the alt text format. We hired a professional sound designer to create background audio for the audio memes, instead of picking from a sound effect library. In future work we would want to additionally evaluate the memes created by novice users.

Study Procedure

Each participant completed a tutorial, listening to the same meme in each format using the screen reader or playing the audio file for the audio macro meme. Then, they were assigned an ordering of the media conditions which were balanced across participants. The meme types (see Appendix A) were randomized for each condition, and examples within each set of five examples were also randomized. They listened to all 5 examples of one meme type, then were asked two questions:

ID	Age	Gender	SM years	Level of vision	Level of vision years	Screen reader
P1	41	M	12	None	10	NVDA
P2	23	M	12	Peripheral, 2 percent central	2	Voiceover, NVDA
P3	53	M	10	None	52	Voiceover, NVDA
P4	45	M	14	None	45	Voiceover, Jaws, NVDA, Narrator
P5	19	M	7	None	19	Voiceover, NVDA
P6	25	F	4.5	None	25	Jaws, NVDA
P7	32	M	12	None	32	Jaws, Voiceover, NVDA
P9	22	F	6	Low vision to total blindness (fluctuates)	19	Voiceover, NVDA, Talkback
P10	19	M	6	Light perception	19	NVDA
P11	39	F	11	None	39	Voiceover

Table 1. Demographics of participants who participated in the online study including age, gender, years on social media (SM years), level of vision, screen reader, and years at the designated level of vision (level of vision years). Note that P8 was unable to complete the study and is excluded here.

1. To what extent do you agree with the statement "I feel I understood the meme" where 1 is Strongly Disagree, 3 is Neutral, and 5 is Strongly Agree?
2. Please describe the meme template (*i.e.* common joke format) to us.

After answering these questions, they completed the same task for two sets of 5 more examples. After completing all 3 meme types for that format condition, they completed the other two conditions. In total, the participants experienced 45 meme examples from 9 meme types. They answered the questions above for each meme type.

Results

The first question posed above seeks to measure the participant's confidence in their understanding of the common joke format for 5 examples of the same meme. We present the average response for each media format by participant in Table 2. Participants were more confident with alt text memes (mean = 3.95), and confidence levels for the text-only (mean = 3.55) and audio macro (mean = 3.52) media formats were similar.

ID	Text Only	Alt Text	Audio Macro	All Conditions
P1	2.67	3.67	3.33	3.22
P2	3.33	4.00	4.67	4.00
P3	4.83	3.83	3.67	4.11
P4	5.00	4.00	3.00	4.00
P5	2.33	2.33	2.83	2.50
P6	5.00	4.67	4.67	4.78
P7	1.33	3.00	1.00	1.78
P9	4.33	5.00	4.67	4.67
P10	2.33	5.00	4.00	3.78
P11	4.33	4.00	3.33	3.89
All	3.55	**3.95**	3.52	3.65

Table 2. The average agreement with "I feel I understood this meme." for each participant by meme format.

The second question we asked after each 5 meme examples was to measure the participants' accuracy of understanding the joke format. Three members of the research team individually wrote the target joke formats, extracting the common elements important to the joke across all of the visual meme examples. These three interpretations of the joke format were combined into a rubric for each example. Two members of the research team redundantly coded a random subset of 20 participant meme templates as either correct or incorrect, and inter-rater reliability was estimated using Cohen's kappa $= 0.7$, which can be interpreted as substantial agreement [17]. One of the team members continued to rate the remaining participant templates. Participant answers were marked correct if they partially or fully matched that meme's rubric, or if they mentioned the name of the meme directly. For example, the rubric for the Success Kid meme was "Victory/outcome/success (especially minor)", and a participant's response of "Little triumphs, little minute triumphs" was rated correct, while "Something bad and then something good." was not specific enough to the form described in the rubric and marked incorrect.

Overall, participants accurately stated 63% of the joke formats after hearing 5 examples in various media conditions. The results across conditions were close, with audio memes having an accuracy of 70%, alt text memes an accuracy of 63%, and text-only memes an accuracy of 57%. Due to the small number of participants, it was not appropriate to perform a statistical analysis on these results, but a larger follow-up study may be able to examine if there is a statistically significant difference between media formats.

Post-Study Interviews

We interviewed each participant about the memes and media formats they experienced after they finished listening to all 45 examples and answering the questions above. Here, we summarize some of their responses and the trade-offs between the different formats.

Format Preferences

The overwhelming majority of participants (8 of 10) preferred the alternative text memes, primarily because it gave them access to a visual description of the content. Several participants noted that this description helped them understand the meme better, particularly if the emotions or facial expressions of the character in the meme were described. Participants often called these "characters" and believed they might be the "speaker" of the meme text. As P3 said regarding the First World Problems meme:

> *It gives you "head in hands, crying". I could get the emotion, but the reason for the emotion appears in the text. – P3*

On the other hand, many participants noted that the images were not always clearly connected with the meme template, and they were confused why it was included.

> *It's a little confusing, because I'm like "Why is a bear saying this?" or "Why is a penguin saying this?" – P6*

This sometimes lead participants to be overly specific about the joke format, such as "Ways the toddler is prevailing over life." for Success Kid, even though a meme example was parking a car, which is an activity not performed by most toddlers.

Participants raised specific concerns about the audio meme format, as it did not use the standard accessibility features (*i.e.* alternative text). This meant the participants did not hear the memes in their preferred voice and speed. Additionally, one participant noted that audio memes are not universally accessible, whereas alternative text or text only memes are available to deaf-blind users or those who use Braille displays.

The participants who preferred formats other than alt text (P6, P9) also reported the most in-depth meme experience in the pre-interview. P6 and P9 noted they found formats other than alt text to be more efficient. While P9 preferred audio memes because the audio quickly conveyed the meme tone (*e.g.*, "dark memes", "sarcasm"), P6 preferred text alone.

Willingness to Share and Create Memes
As many of the participants had not experienced a large number of internet memes before, we asked them if they would have posted any of the 45 examples they experienced during the study. Nine of the participants had at least one they might post, but several would only do so with friends, not publicly. P9 was very enthusiastic about sharing memes in general – just not the ones we chose as examples:

> *I would probably consider posting them because they were strictly made in an accessible format, [But] my friends would think "Why are you posting things from 2011?" – P9*

Three participants said they would definitely create memes themselves if they had tools to do so.

> *I certainly want to be part of the culture. There are circumstances where I think the message I am trying to convey would be done better by visual memes than verbal or writing. It's so easy and it's so efficient to share when a picture can convey a message. – P1*

Three participants were not confident they would be able to create memes without sight, as the visual component is important. Four participants stated they were not interested in creating memes themselves, but would like to view them.

DISCUSSION
Our interviews and user studies with the ten Twitter users with vision impairments revealed a number of opinions and preferences about meme media formats.

Primarily, the users sought access to the same information provided to sighted users: a description of the visual image and the overlaid text. In some cases this helped the participants understand the humor or other sentiment in the meme (*e.g.*, First World Problem), although in a few cases it was confusing (*e.g.*, Confession Bear). The users stated the audio and text memes did not provide enough context to understand the meme, and this is reflected in their confidence ratings for these conditions. However, the users had similar accuracy scores for memes in these conditions, indicating there might be a divide between confidence and actual understanding of the different formats.

Some of the stated concerns with the audio memes may be due to its unfamiliarity. They were not integrated with screen readers, so they did not automatically play on focus like the alternative text. They also did not use preferred voices or speaking rates. Close integration with screen readers could alleviate these problems with audio memes, but other issues, such as lack of universal accessibility, are inherent to the media format. As the system can produce text-only, alt text, and audio memes, we can create accessible content in multi-modal formats, allowing users to select their preferred formats.

We followed established guidelines for creating meme alt text [10, 26]. Still, our alt text did not always highlight information users needed to understand memes. Specifically, users requested more information about the character in the meme and their emotional state. In addition, several users mistook the image style of memes when reporting what they imagined the meme to look like (*e.g.*, reporting the images to be low-effort drawings or stick figures instead of photographs). Based on prior work [26] and our study results, we propose a condensed, meme specific set of structured questions for writing alt text of memes:

- Who are the character(s) in these memes?
- What actions are the characters performing, if any?
- What emotions or facial expressions do the character(s) exhibit in these examples?
- Do you recognize the source of the image (TV show, movie, etc)? If so, what is it?
- Is there anything notable, or different about the background of the image?

Meme descriptions that provide this type of context remain consistent with the fact that much of the humorous effect comes from a character acting out a scenario rather than simply describe it [14, 31]. By describing who is acting out the meme text, and what the image indicates about their background, we may be able to give viewers the intended experience.

Limitations and Future Work
In the user study with Twitter users with vision impairments, we presented meme examples that were crafted by members of the research team. These examples represent some of the best case scenarios for each format. Word errors in the OCR results were corrected, alt text was written with best practices in mind [26, 10], and the background audio in the audio memes were created by a professional sound designer. Online volunteers or crowd workers may not generate alternative meme templates of the same quality, although prior work demonstrates that this is true in the case of alternative text [26].

We operated from a known set of historical memes curated by Imgur and Meme Generator, but in reality new memes are always being created or modified. These examples may not exist in our database, or they may be similar enough to another meme to match, but have a different semantic meaning. Future work should explore how quickly a new meme in the wild can be recognized, and how many examples of the meme are needed before it can be transformed into an accessible format.

Internet memes are so commonly associated with visual content that most participants did not imagine audio memes beyond accessible versions of images. We believe that memes generated as audio first by people with vision impairments may be interesting as a standalone non-visual media, especially for other blind users. This may open up opportunities to explore multi-modal representations of memes and online content. In addition to static memes, participants mentioned they would like access to GIFs that are commonly posted on Twitter as reactions to tweets. Audio descriptions of GIFs could be similar to those provided for accessible videos.

CONCLUSION

Memes may not always be vehicles for conveying serious content, but they remain an important part of online discourse, whether that is public or in small groups with friends. Creators of memes typically do not include alternative text, rendering almost all of them inaccessible to people with vision impairments. We have presented an automatic method to recognize known memes, extracting the overlaid text, and rendering that text into a more accessible format, such as alternative text or an audio meme template. Because many memes are repeated images with new text, this results in a scalable solution to make a large number of online memes accessible just by creating alternative text or audio versions of the base meme template.

In a study with 10 Twitter users with vision impairments, we found that they preferred the alternative text memes due to their inclusion of visual context, compatibility with screen readers, and universal accessibility. The study also reveals that people with vision impairments are eager to share accessible memes, as they are a part of culture and communication online. Based on their responses, we propose a short set of structured questions for alternative text authors to answer when describing memes. These can assist the authors using our system to not only make memes trivially accessible, but also preserve the emotional tone or humor embedded in the meme. Even the participants who were not as interested in "silly" memes noted that their lack of alternative text was a source of significant accessibility issues on social media.

> *I think [memes] could become a way to generate a lot of useless content very quickly. But if there has to be a lot of useless content out there, it ought to be accessible. – P4*

ACKNOWLEDGEMENTS

This work has been supported by the National Science Foundation (#IIS-1816012 and #DGE-1745016), Google, and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR 90DP0061).

REFERENCES

[1] Tim Berners-Lee and Dan Connolly. 1995. *Hypertext Markup Language - 2.0*. RFC 1866. Internet Engineering Task Force. 1–77 pages. DOI: http://dx.doi.org/10.17487/RFC1866

[2] Jeffrey P. Bigham, Ryan S. Kaminsky, Richard E. Ladner, Oscar M. Danielsson, and Gordon L. Hempton. 2006. WebInSight:: Making Web Images Accessible. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '06)*. ACM, New York, NY, USA, 181–188. DOI: http://dx.doi.org/10.1145/1168987.1169018

[3] Lei Chen and Chong Min Lee. 2017. Convolutional Neural Network for Humor Recognition. *CoRR* abs/1702.02584 (2017).

[4] W3 Consortium. 2018. Web Content Accesisbility Guidelines (WCAG) 2.1. (2018). https://www.w3.org/TR/WCAG21/

[5] W3 Consortium and others. 1998. *HTML 4.0 specification*. Technical Report. Technical report, W3 Consortium, 1998. http://www.w3.org/TR/REC-html40

[6] Daniel Dardailler. 1997. The ALT-server ("An eye for an alt"). (1997).

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of 2009 Computer Vision and Pattern Recognition (CVPR '09).*

[8] Marta Dynel. 2016. "I has seen Image Macros!" Advice Animals memes as visual-verbal jokes. *International Journal of Communication* 10 (2016), 29.

[9] Meme Generator. 2019. The Most Popular Memes of All Time. (2019). https://memegenerator.net/memes/popular/alltime

[10] Cole Gleason, Patrick Carrington, Cameron Cassidy, Meredith Ringel Morris, Kris M. Kitani, and Jeffrey P. Bigham. 2019. "It's Almost Like They'Re Trying to Hide It": How User-Provided Image Descriptions Have Failed to Make Twitter Accessible. In *The World Wide Web Conference (WWW '19)*. ACM, New York, NY, USA, 549–559. DOI: http://dx.doi.org/10.1145/3308558.3313605

[11] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions Using Reverse Image Search. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 518, 11 pages. DOI: http://dx.doi.org/10.1145/3173574.3174092

[12] Stephanie Hackett, Bambang Parmanto, and Xiaoming Zeng. 2004. Accessibility of Internet Websites Through Time. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '04)*. ACM, New York, NY, USA, 32–39. DOI: http://dx.doi.org/10.1145/1028630.1028638

[13] Xiaodong He and Li Deng. 2017. Deep Learning for Image-to-Text Generation: A Technical Overview. *IEEE Signal Processing Magazine* 34, 6 (nov 2017), 109–116. DOI:http://dx.doi.org/10.1109/MSP.2017.2741510

[14] Sally Holloway. 2010. *The Serious Guide to Joke Writing: How To Say Something Funny About Anything*. Bookshaker, Great Yarmouth, UK. 207 pages.

[15] Instagram. 2018. Creating a More Accessible Instagram. (2018). https://instagram-press.com/blog/2018/11/28/creating-a-more-accessible-instagram/

[16] Chloé Kiddon and Yuriy Brun. 2011. That's What She Said: Double Entendre Identification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2 (HLT '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 89–94. http://dl.acm.org/citation.cfm?id=2002736.2002756

[17] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[18] Chi-Chin Lin, Yi-Ching Huang, and Jane Yung jen Hsu. 2014. Crowdsourced Explanations for Humorous Internet Memes Based on Linguistic Theories. In *Proceedings of AAAI Conference on Human Computation and Crowdsourcing (HCOMP '14)*.

[19] Daniel C. Dennett Matthew M. Hurley. 2011. *Inside Jokes: Using Humor to Reverse-Engineer the Mind*. The MIT Press, Cambridge, MA, USA. 376 pages.

[20] Microsoft. 2017. Seeing AI | Talking camera app for those with a visual impairment. (2017). https://www.microsoft.com/en-us/seeing-ai/

[21] Rada Mihalcea and Carlo Strapparava. 2005. Making Computers Laugh: Investigations in Automatic Humor Recognition. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 531–538. DOI: http://dx.doi.org/10.3115/1220575.1220642

[22] 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 (CHI '18)*. ACM, New York, NY, USA, Article 59, 11 pages. DOI: http://dx.doi.org/10.1145/3173574.3173633

[23] Library of Congress. 2012. Homepage | Meme Generator. (2012). https://www.loc.gov/item/lcwaN0010226/

[24] Victor Raskin. 2009. *The Primer of Humor Research*. De Gruyter, Berlin, Germany. 673 pages.

[25] Adrian Rosebrock. 2014. The complete guide to building an image search engine with Python and OpenCV. (Dec 2014). https://www.pyimagesearch.com/2014/12/01/complete-guide-building-image-search-engine-python-opencv/

[26] Elliot Salisbury, Ece Kamar, and Meredith Ringel Morris. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. In *Proceedings of 2017 AAAI Conference on Human Computation and Crowdsourcing (HCOMP '17)*. 147–156. www.aaai.orghttps://www.microsoft.com/en-us/research/wp-content/uploads/2017/08/scalable

[27] Dafna Shahaf, Eric Horvitz, and Robert Mankoff. 2015. Inside Jokes: Identifying Humorous Cartoon Captions. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. ACM, New York, NY, USA, 1065–1074. DOI:http://dx.doi.org/10.1145/2783258.2783388

[28] Hironobu Takagi, Shinya Kawanaka, Masatomo Kobayashi, Takashi Itoh, and Chieko Asakawa. 2008. Social Accessibility: Achieving Accessibility Through Collaborative Metadata Authoring. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '08)*. ACM, New York, NY, USA, 193–200. DOI: http://dx.doi.org/10.1145/1414471.1414507

[29] Julia M. Taylor and Lawrence J. Mazlack. 2004. Computationally recognizing wordplay in jokes. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci '04)*, Vol. 26.

[30] Twitter. 2018. How to make images accessible for people. (2018). https://help.twitter.com/en/using-twitter/picture-descriptions

[31] John Vorhaus. 1994. *The Comic Toolbox: How to Be Funny Even If You're Not*. Silman James Press, Los Angeles, CA. 191 pages.

[32] Ye-Yi Wang, Alex Acero, and Ciprian Chelba. 2003a. Is Word Error Rate a Good Indicator for Spoken Language Understanding Accuracy. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '03)*. IEEE, 577–582.

[33] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003b. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Vol. 2. Ieee, 1398–1402.

[34] Elwyn Brooks White. 1954. Some remarks on humor. *The Second Tree from the Corner* (1954), 173–181.

[35] 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 (CSCW '17)*. ACM, New York, NY, USA, 1180–1192. DOI: http://dx.doi.org/10.1145/2998181.2998364

Figure 6. An example of each meme template. In the study, we used 5 example memes for each meme template for 45 total memes.

APPENDIX

A: MEME TEMPLATES

In our study, we used nine different visual memes (Figure 6) with five examples for each. The names of the memes we used, are listed here:

A Awesome Awkward Penguin
B Success Kid
C Philosoraptor
D Bad Luck Brian
E Most Interesting Man in the World
F Confession Bear
G Awkward Moment Seal
H First World Problems
I Futurama Fry

We include the alt text template for each meme (Table 3) and a meme example for each (Figure 6).

Base meme	Alt text template
Confession Bear	Baby black bear staring into space with paws on a tree branch. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Success Kid	Toddler clenching fist in front of a smug face. Overlaid text on top [top text]. Overlaid text on bottom [bottom text]
Awkward Moment Seal	Close up of a seal's face with wide eyes and a straight face. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Interesting Man	A man with gray hair in a nice shirt and jacket smirking while leaning on one elbow. A bottle of Dos Equis beet is in front of him. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Philosoraptor	A drawing of a green dinosaur raptor with a claw to it's chin and mouth open as if it is contemplating something. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
First World Problems	Close up on a woman with her eyes closed head in one hand and a stream of tears running down her cheek. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Awesome Awkward Penguin	Close up of a seal's face with wide eyes and a straight face. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Bad Luck Brian	A young kid in an awkward school photo. He is wearing a plaid vest and has an open smile where you can see his braces. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].
Futurama Fry	Fry from the show Futurama a cartoon man with orange hair squinting his eyes as if he suspects something. Overlaid text on top [top text]. Overlaid text on bottom [bottom text].

Table 3. Names of memes (base meme) with the corresponding alt text template for each meme. When the template lists [top text] and [bottom text], we replace the placeholders with the example meme text. Audio meme templates included in the supplemental materials.