BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments

Abigale J. Stangl*[1], Esha Kothari*[2], Suyog D. Jain[3], Tom Yeh[1], Kristen Grauman[2,4], Danna Gurari[2]

* denotes equal contribution

[1] University of Colorado Boulder, [2] University of Texas at Austin, [3] CognitiveScale, [4] Facebook AI Research

ABSTRACT

Our interviews with people who have visual impairments show clothes shopping is an important activity in their lives. Unfortunately, clothes shopping web sites remain largely inaccessible. We propose design recommendations to address online accessibility issues reported by visually impaired study participants and an implementation, which we call BrowseWithMe, to address these issues. BrowseWithMe employs artificial intelligence to automatically convert a product web page into a structured representation that enables a user to interactively ask the BrowseWithMe system what the user wants to learn about a product (e.g., What is the price? Can I see a magnified image of the pants?). This enables people to be active solicitors of the specific information they are seeking rather than passive listeners of unparsed information. Experiments demonstrate BrowseWithMe can make online clothes shopping more accessible and produce accurate image descriptions.

INTRODUCTION

Shopping online rather than in a brick-and-mortar store can be beneficial for a variety of reasons. Time can be saved and scheduling difficulties can be avoided by not having to travel to and from the store in the restricted set of hours the store is open. Hassles of finding the location of the article of interest within the constantly changing layout of a store (if it is even available in the appropriate size) can be avoided. In addition, shopping online can often be cheaper since a person can quickly compare prices of the same item from different manufacturers before choosing where to buy.

Yet, people with visual impairments (i.e., low vision and blind) who were asked "What are a few web sites or types of web sites that you would like to visit, but avoid because of accessibility issues?" responded that flash-based web sites and shopping web sites were rated as the most problematic (the top 2) [2]. While encouraging progress has been made to address the top challenge through crowdsourcing [33], little effort has been made to understand why online shopping is inaccessible and how to address it.

Our goal is to improve the online shopping experience for people with visual impairments. Figure 1 exemplifies a typical

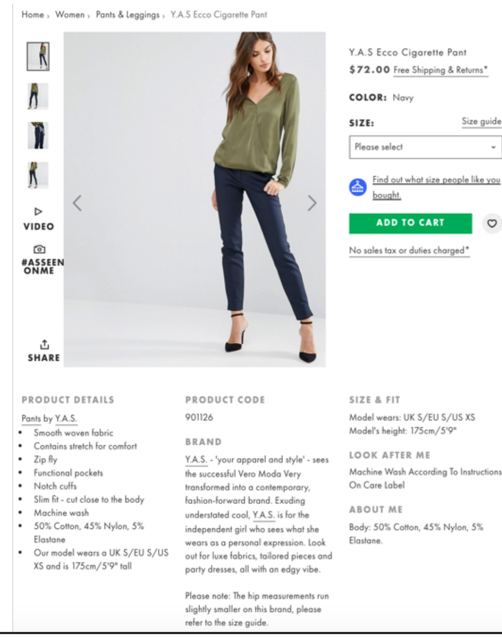
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASSETS '18, October 22–24, 2018, Galway, Ireland

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-5650-3/18/10. . . 15.00

DOI: https://doi.org/10.1145/3234695.3236337



Screen Reader Output: [~205 words spoken followed by this product description]: heading level 1, Y.A.S ecco cigarette pant, $72.00, internal, link, free shipping and returns, color, navy, size, internal, link, size guide, please select, collapsed, pop up button, XL-US 10 menu item, L-US 8 menu item, M-US 6 menu item, S-US 6 menu item, XS-US 2 menu item, Please select, menu item, link, find out what size people you like bought, link add to cart, internal, link, 6987053, link, No sales or duties charged, heading level 4, product details, link Pants, by, link, Y.A.S, list 9 items, Smooth woven fabric, contains stretch for comfort, zip fly, functional pockets, notch cuffs, slim fit-cut close to the body, machine wash, 50% cotton, 45% nylon, 5% elastane, our model wears UK S/EU S/US XS and is 175cm/5'9" tall, heading level 4, product code, 901126, heading level 4, brand, link, Y.A.S, -'your apparel and style' - sees the successful very moda very transformed into a contemporary, fashion-forward brand, exuding understated cool, link, Y.A.S, is for the independent girl who sees what she wears as a personal expression, look out for luxe fabrics, tailored pieces and party dresses, all with an edgy vibe, please note: the hip measurements run slightly smaller on this brand, please refer to the size guide, heading level 4, size and fit, model wears UK S/EU S/US XS, model's height 175cm/5'9", heading level 4, look after me, machine wash according to instructions on care label, heading level 4, About me, Body, 50%cotton, 45% Nylon, 5% Elastane.

Alt Text: [None provided]

Figure 1. Often shopping web sites are multimodal, describing each product with text and images. As shown for ASOS.com, the information read by a screen reader is painstakingly lengthy and the image description (i.e., Alt text) is missing. Our aim is to improve the accessibility of shopping web sites. We design, prototype, and evaluate a system called BrowseWithMe that enables a user to actively query a web page for the specific multimodal information the user is seeking.

product web page and offers a few clues as to why shopping web sites may be inaccessible—information overload and inadequate image description. To access content about the product, a person is required to listen to a pre-defined, one-size-fits-all list of the web page contents to locate the pertinent information (e.g., price?). In addition, access to images is constrained due to inadequate or missing text descriptions (i.e., Alt text).

Towards our goal, we focus on *clothes shopping* as a non-trivial, real-world example. We first conduct an empirical investigation to examine the factors that impact this population's clothes shopping experiences in brick-and-mortar stores and online. Our findings highlight that participants want to shop for clothes online, but often are inhibited by the widespread inaccessibility of web sites. Congruent with these findings, we identify design factors to address online accessibility issues and develop a prototype implementing these for a key stage of the shopper's journey—learning about a product described on a web page (e.g., Figure 1). We call this system BrowseWithMe. BrowseWithMe empowers a user to navigate a product web page at one's own pace by asking about and receiving only the desired information; e.g., "what material is the pants?" or "show me a magnified image of the pants." Experiments demonstrate our proof-of-concept implementation of BrowseWithMe yields more accurate image descriptions than Alt text and improves the online shopping experience.

To our knowledge, BrowseWithMe is the first automated assistant that enables visually impaired users to actively ask for and receive only the information they are seeking on a web page. This is in contrast to existing screen readers which often deliver a lengthy list of all web page contents, as illustrated in Figure 1. The key technical aspect behind our design of an automated, interactive online shopping assistant is an artificial intelligence back-end that converts multimodal input data (e.g., text and images) into a structured representation.

RELATED WORKS

Shopping in Brick-and-Mortar Stores. Our work relates to a body of literature aimed at understanding and addressing challenges people with visual impairments encounter when shopping in brick-and-mortar stores. For example, [44] reports that "finding a product in the store was the most challenging subtask" for people with low vision and [46] reports that blind people "often relied on sighted guides when visiting the mall and other stores, as these locations rarely offered accessible maps". Accordingly, numerous technological solutions have been proposed to improve the shopping experience in brick-and-mortar stores, such as bar code readers [25], RFIDs [29], robotics [19], computer vision based systems [47, 52], and braille product labels [6]. Our work complements existing work by contributing richer information about the habits, motivations, and challenges people with visual impairments face when shopping in brick-and-mortar stores in the domain of clothes shopping. Unlike prior work, our work is also situated within the broader landscape of web accessibility. Specifically, we also examine the experiences of people with visual impairments when shopping for clothes online and propose a technology to improve the online shopping experience.

General Web Accessibility. Our work also relates to the literature that aims to understand and address web accessibility challenges for people with visual impairments. One body of work focuses on the Web Content Accessibility Guidelines (i.e., WCAG), the most comprehensive protocol for developing accessible web sites. For example, some studies demonstrate existing web pages inadequately implement the guidelines [12, 21, 37] while others demonstrate that meeting WCAG guidelines does not guarantee fewer accessibility challenges [38]. Accordingly, several works propose alternative solutions, such as empowering web users to flag accessibility issues for a human-powered service to fix [23] and re-narrating web sites to communicate only a select set of information chosen by a user a priori [16, 39]. Our work complements prior work by contributing richer information about the online accessibility challenges faced in the domain of clothes shopping. Furthermore, we offer a novel design for overcoming online accessibility challenges which relies on artificial intelligence as a backend that responds to users' requests to learn about multimodal data (e.g., text and images) on web pages. This is a promising alternative to relying on developers to follow WCAG guidelines or users to supervise the shared content.

Image Accessibility. Numerous solutions have been proposed to overcome the well-known challenge that images are inaccessible to people with visual impairments. For example, the pioneering ESP game [5] was an important catalyst for employing crowd workers to generate descriptions of images [15, 24, 40, 45]. Subsequently, numerous algorithms [11, 48] were proposed to automatically generate the image descriptions more efficiently and cheaply. More recently, methods are emerging that enable a user to interactively explore various objects in a 2D image at his/her own pace by both localizing objects in an image and attaching a description to each object [27, 30, 51, 53]. While systems have been proposed for interactive image exploration in the problem domain of online shopping [27, 30, 32, 51], to our knowledge no prior work has studied its value for people with visual impairments. Our studies reveal the unique interests of this population and design opportunities for improving the accessibility of online shopping.

Magnification Tools. Tools that enable users to enlarge the content of interest on the screens of their personal technological devices are highly valued by people with low vision [3, 17]. Such tools enlarge specific elements using a variety of cues, such as the content underneath a mouse [31], a person's physical distance from the screen [22], the location pointed at with a person's fingertip [42], or web page elements automatically detected to occupy unnecessarily large amounts of white space [9]. Unlike prior work, our approach relies only on spoken (or typed) user requests to automatically identify which items in an image to magnify. Our studies reveal an excitement from individuals with low vision about this efficient, user-centric magnification approach for images.

On-Demand Personal Assistants. Many systems offer on-demand assistance to people with visual impairments. For example, human-powered assistants such as VizWiz [10], Be My Eyes [1], and Legion:View [33] empower its users to learn answers to their visual questions from sighted remote helpers. Automated assistants such as Apple's Siri, Google Now, and Amazon's Alexa respond to voice commands to execute simple tasks such as performing web search, updating calendars with new events, or purchasing items. While the aforementioned assistants were developed for general-purpose applications, we argue that special design considerations must be made to better meet the needs of a specific user population (blind and low vision) for a specific application (online clothes

ID: Alias	Age/Gender	Visual Impairment: Cause (Age of Diagnosis Onset)	Technology Preference
P1: Jackie	37/F	Blind (total): Destroyed optic nerve (34)	iPhone/Voice Over
P2: Lily	56/F	Blind (total): Retinitis (40)	iPhone/Voice Over
P3: Bob	28/M	Blind (total): Peter's anomaly (birth)	PC/NVDA
P4: Annie	34/F	Blind (some light perception): Acquired (14)	PC/NVDA
P5: Sam	26/M	Low vision: Stargardt's disease (birth)	PC/Magnification
P6: Cathy	36/F	Low vision: Undisclosed (birth)	iPhone/Magnification
P7: Julia	39/F	Low vision: Macular atrophy/degeneration (32)	iPad/Magnification & Voice Over
P8: Vivian	56/F	Low vision: Glaucoma (46)	iPhone/Magnification & Voice Over

Table 1. Demographics and technology preferences of participants in our experimental investigation (and subsequent user study).

shopping). Our work contributes knowledge about such needs and a potential solution to address them.

Choosing Clothes. Prior work has offered solutions to assist people with visual impairments to choose clothes to wear. For example, some systems solicit anonymous lay people to describe articles of clothing (e.g., clothing color) [8, 13] and offer subjective fashion advice [14] while other systems automatically advise which clothes in a wardrobe match [36, 50]. Our work fills a gap in the literature by enriching our understanding of an earlier step in the process of choosing clothes to wear—the task of acquiring (shopping) for clothes.

EMPIRICAL INVESTIGATION

Our first aim is to better understand how people with visual impairments experience clothes shopping. We conducted our analysis in the broader context of brick-and-mortar and online shopping experiences to inform our effort to make online clothes shopping experiences more accessible.

Interview Design

We conducted semi-structured interviews, guided by Value Sensitive Design (VSD) [18], with eight study participants with visual impairments. We began by asking participants 20 open-ended questions to learn about the significance they place on clothing shopping in their lives, their experiences and habits when shopping for clothes online versus in brick-and-mortar stores, and their use of technology and other assistance while clothes shopping. We also asked 13 Likert-scale questions focused on the participants' value and expected value of clothes shopping and how accessibility factors impact their shopping experiences (1=Very low / 5=Very high). Finally, we asked participants to demonstrate how they approach online shopping using their preferred technology. Each interview lasted from 30 to 45 minutes, and took place in a location chosen by the participant; we met four participants in their homes, two participants in coffee shops, and two participants in a university office.

We then analyzed the data to identify emergent trends. To support data analysis, we audio and video recorded each interview. We then transcribed each audio recording and performed open coding [43] on the transcriptions to identify trends.

Table 1 summarizes demographics for all participants (female: 8, male: 2). The age of the participants ranges from 26 to 58 (Mean=36.5, SD=11.38). Two participants described themselves as totally blind, two participants described themselves

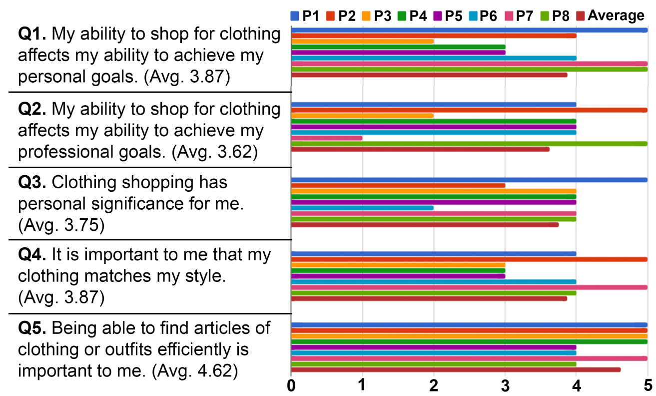


Figure 2. Answers to five Likert scale questions asked about general clothes shopping experiences; 1= Strongly Disagree; 2= Disagree; 3= Neutral; 4=Agree; 5=Strongly Agree.

as blind with very limited light perception, and four participants reported having low vision. We recruited using connections with a local university, a local independence training center, and snowball sampling. We did not recruit for a specific type of visual impairment as our aim was to understand the range of experiences faced by people with various visual impairments. All participants completed the interview and were compensated with a 20 USD Amazon Gift Card (for participation in this inquiry and our subsequent user study).

Importance of Clothes Shopping

Overall, our Likert scale data shows clothes shopping is personally significant for the participants; i.e., 6 of the 8 participants agree or strongly agree (i.e., Q3; Figure 2). Most participants reported that clothes shopping impacted their ability to achieve their personal goals (i.e., average of 3.87/5; Q1; Figure 2) and professional goals (i.e., average of 3.62/5; Q2; Figure 2). This is interesting given that participants brought such a large diversity of experiences with shopping, stemming from their unique interests and experiences with vision loss (e.g., Table 1). Our understanding of these numbers is enriched by the following three emergent themes from the interviews which suggest why clothes shopping is important to them:

- **Clothing is necessary**: As articulated by Bob in jest, "*Nothing makes clothing shopping necessary; it's necessary because they do not let us run around naked yet.*"
- **Clothes shopping is a social activity**: For Annie, Sam, and Cathy, their social experiences were enhanced by receiving

trusted feedback about fit, style, etc. as well as sharing in the joy of finding quality items and good deals.

- **Clothes shopping is an activity that supports expressing personal interests and identity**: For Sam, he expressed a strong need to shop for specialty shirts online that "*show my creativity and aesthetic preferences.*" Lily and Julia noted their professional identity is related to shopping, due to the need to both perform shopping and manage stores. More broadly, the Likert scale data shows that most participants (i.e., 5/8) agree or strongly agree that having their clothing match by style is important (i.e., Figure 2; Q4).

Current Experiences with Clothes Shopping

All eight participants shared they most often shop for clothes in brick-and-mortar stores, and five participants reported experience with clothes shopping online. We here discuss the factors that they reported affect their ease and difficulty with shopping in brick-and-mortar stores and online. From open-coding analysis, we identified three key factors: 1) accessing product information; 2) use of shopping assistance; and 3) achieving personal mobility, orientation, and transportation. We summarize our findings with respect to these factors below, both for shopping in brick-and-mortar stores and online.

Obtaining Product Information

We found locating accurate information is a critical factor behind our participants' satisfaction with a shopping experience.

With respect to shopping in brick-and-mortar stores, all study participants reflected that the biggest benefit, as Julia put it,"*is the ability to know what I am buying and getting what I am seeking and knowing that I am not getting ripped of in terms of price or quality.*" However, as noted by one participant, difficulties include the inability to read price tags and size information on clothing items.

With respect to online shopping, all participants noted they could potentially get more information from web-based product descriptions. Three participants mentioned the ease in which shopping web sites like Amazon make it to read reviews to obtain extra information. However, online shopping currently eliminates the ability to touch products. Furthermore, it presents significant challenges due to many web sites failing to adhere to WCAG Standards. Cindy, Jackie, and Vivian all described abandoning online shopping because voice over often does not work or images lack Alt Text. Several participants with low vision (Julia, Scott, and Vivian) also noted challenges while trying to access visual content despite their ability to use magnification tools on their devices. Vivian said, "*Sometimes they [the images] are OK, sometimes they are too small or they do not give a full view.*" Jackie noted a difficulty with using existing accessibility technologies is that they require a lot of mental effort to parse and recall all the information that "*comes at me.*" Limited fluency with accessibility technologies (e.g., screen-readers) also impacted some participants.

Shopping Assistants

Many participants shared their satisfaction with a shopping experience is impacted by the type(s) of assistance they receive. With respect to brick-and-mortar experiences, some stores offer services to facilitate shopping experiences. Three participants referenced positive experiences with assistants who directly approached them and asked if they need assistance. As Vivian noted, when assistants are informed and considerate they can create positive experiences. But, she noted the responsibility largely falls on the customer to "*try to give them as much as possible; the occasion [for wearing the item] and the style*". Similarly, Annie shared, "*I only use somebody at a store for help when I am looking for something very specific because they do not know my standards.*" Lily reflected, "*[A] few times I have got really great people, but very rare.*" Five participants shared negative experiences with shopping assistants. Jackie described feeling socially isolated when assistants do not approach her. Lily and Jackie also expressed concern over how well the shopping assistants are trained to listen to them or know enough about the products they are trying to sell. Julia also expressed a lack of confidence in shopping assistants for fear that they are "*trying to get a commission out of me.*" Cathy described "*They always take me all over the place to try to find what we are looking for but we do not find it.*" Despite an appreciation that brick-and-mortar shopping assistants exist, collectively the participants shared shopping assistants are only helpful when they are patient, respectful, attuned to fashion, trained in giving verbal directions, and help them find alternative sizing, colors, and materials.

With respect to online shopping, none of the participants had experience with an on-demand personal assistant. However, several participants raved how online product reviews assist their desire for more product information. According to Julia, "*I rely on reviews a lot if I shop online. I also rely on the ability to leave questions. When I leave reviews, I get questions that people post online. I find that comfortable to create the conversation and be honest about what the product is and is not.*" A reoccurring concern about using on-demand personal assistants was the fidelity or trustworthiness of information. Comments resembled that from Vivian who shared "*I do not trust web sites like Ebay, or Craigslist. I feel that it is sketchy without at least getting reviews.*"

Mobility/Orientation/Transportation

The final reoccurring theme pertained to a difficulty with brick-and-mortar shopping. Specifically, all participants discussed frustrations with traveling to and navigating through malls and stores, seven of whom explicitly shared about the challenge with navigating to stores. As Vivian explained, "*Clothing shopping is a serious commitment..,it is going to be two to three buses or trains to get to where I need to go.*" Bob shared that even once he gets to the mall, "*it's a nightmare; there is no GPS. If they had that, my phone would have been working inside which probably would have told which store I am inside.*" Similarly, Sam noted "*Having a visual disability makes it [getting around stores] more stressful. I have to use my phone, and I am very close to it. I look retarded.*" Julia noted that many retail stores are designed for the visual experience, in order to help people find outfits rather than to "*easily make their way around the space.*" Moreover, she shared "*Now that I have lost my vision, I am not comfortable with public transportation or ride-share services, and I do not want to impose* *on my husband.*” We found that the time it takes to travel somewhere, the lack of accessible transportation offerings, concern about their personal mobility and orientation skills, the stigma of having to rely on another person for transport, or the challenge of finding their way around a mall or store are major deterrents for shopping in brick-and-mortar stores.

Every participant indicated that online shopping offers great promise to reduce travel time and effort.

Trade-Offs Made When Clothes Shopping

When deciding how to shop for clothing, our participants regularly made trade-offs between the aforementioned factors. For example, when shopping at brick-and-mortar stores, participants negotiated between the desire to learn more about products and the burden of relying on assistants. When shopping online, participants negotiated between the desire to efficiently find comprehensive, trustworthy information about products and the burden of learning a web site's particular conventions and accessibility issues. And when choosing whether to shop in a brick-and-mortar store versus online, participants negotiated between (1) the desire for independence from assistants in brick-and-mortar stores versus the time and effort to develop the digital literacy needed to shop online and (2) the desire to feel and try on clothing at brick-and-mortar stores versus the potential time and effort saved by shopping online (by avoiding mobility, orientation, and transportation challenges).

Discussion

Our findings offer evidence that the target population finds clothes shopping important to achieve personal and professional goals. Moreover, they seek distinct experiences when choosing to shop in brick-and-mortar stores versus online, in part, due to accessibility challenges that impact their ability to shop in each environment. Our findings also underscore that accessibility challenges can vary depending on a person's vision level (i.e., blind versus low vision), for example, whether a web site has a non-functioning magnification tool versus an image with no Alt text. Broadly, we summarize the accessibility challenges affecting participants as revolving around the following values: 1) Efficiency and Ease; 2) Information Access and Trust; 3) Independence; 4) Social Connection; and 5) Freedom of Expression and Personalization.

While our empirical investigation reveals a broad range of opportunities to improve the accessibility of shopping in brick-and-mortar stores and online, we focus for the remainder of this paper on addressing the accessibility of shopping online.

AUTOMATED ONLINE SHOPPING ASSISTANT

In what follows, we describe our design recommendations and prototype for an accessible online shopping assistant.

Design Recommendations

Our empirical observations revealed the types of accessibility challenges could vary based on the participants' level of vision and experience using online shopping web sites. Towards improving online accessibility for blind and low vision users, we propose the following three design recommendations:

Consistent Structure. We repeatedly observed user inefficiency and frustration because of a lack of consistency between web sites. For example, low-vision participants shared a frustration with having to continuously learn how to use different magnification tools that are provided on different web sites. Julia describes her experience using her iPad to browse different web sites: “*All the magnification and enlargement stuff will work on Ebay. When I navigate to the Columbia web site it works. But then if I go through the app, nothing can enlarge [by zooming in on the page]. Then I have to open up the [iPad] magnifier, but it doesn't let me move around [the web page]. Ninety percent of the time when I try a web page's version, I am disappointed in the accessibility compatibility.*” In addition, we heard frustration about web sites that do not follow WCAG guidelines. *Bob*, a computer scientist, asked for developers of online shopping web sites to “*Use CSS properly — don't use CSS to do things that HTML can do. Label images, use buttons that are clearly accessible, and add good descriptions in a succinct manner!*”. **We recommend the design of a single, domain-specific structure which users can then query to address their questions, and the implementation of a web site wrapper that automatically converts any product web page to fit into this structured representation.**

Complete Information. We also observed issues arise due to an incomplete understanding about a product. In some cases, the incomplete understanding occurred because information was missing. For example, participants reported Alt text to describe an image often was missing or inadequate, despite clear WCAG guidelines requiring it. In other cases, the incomplete understanding occurred due to uncertainty where to locate the information of interest. For example, some participants who demonstrated their use of personal devices to shop online would eventually abandon their efforts to answer a question about a product because they questioned whether the information was available on the web page. **While a structured representation of a web page can help a user learn if the information of interest is present, we also recommend using automated tools to fill in information found to be missing when creating the structured representation of a web page (e.g., Alt text for images).**

Hierarchical Structure. Additionally, we repeatedly observed user inefficiency and frustration arise because of a lack of control over the level of detail received. For example, when *Jackie* demonstrated how she used her personal device to shop online, she pleaded for information to be presented “*up front and then let the shopper choose to go into detail.*” In addition, low vision participants wanted to zoom into images to observe features of clothing in greater detail. **We recommend designing a hierarchy for the structured representation so users can learn about a product at their preferred level of detail (e.g., display full image versus magnified view of an item in the image; communicate the product name versus the lengthy product description).**

BrowseWithMe Implementation

Next, we present our system for improving accessibility—an online shopping assistant which possesses a set of skills to meet the design recommendations above. Inspired by the ac-

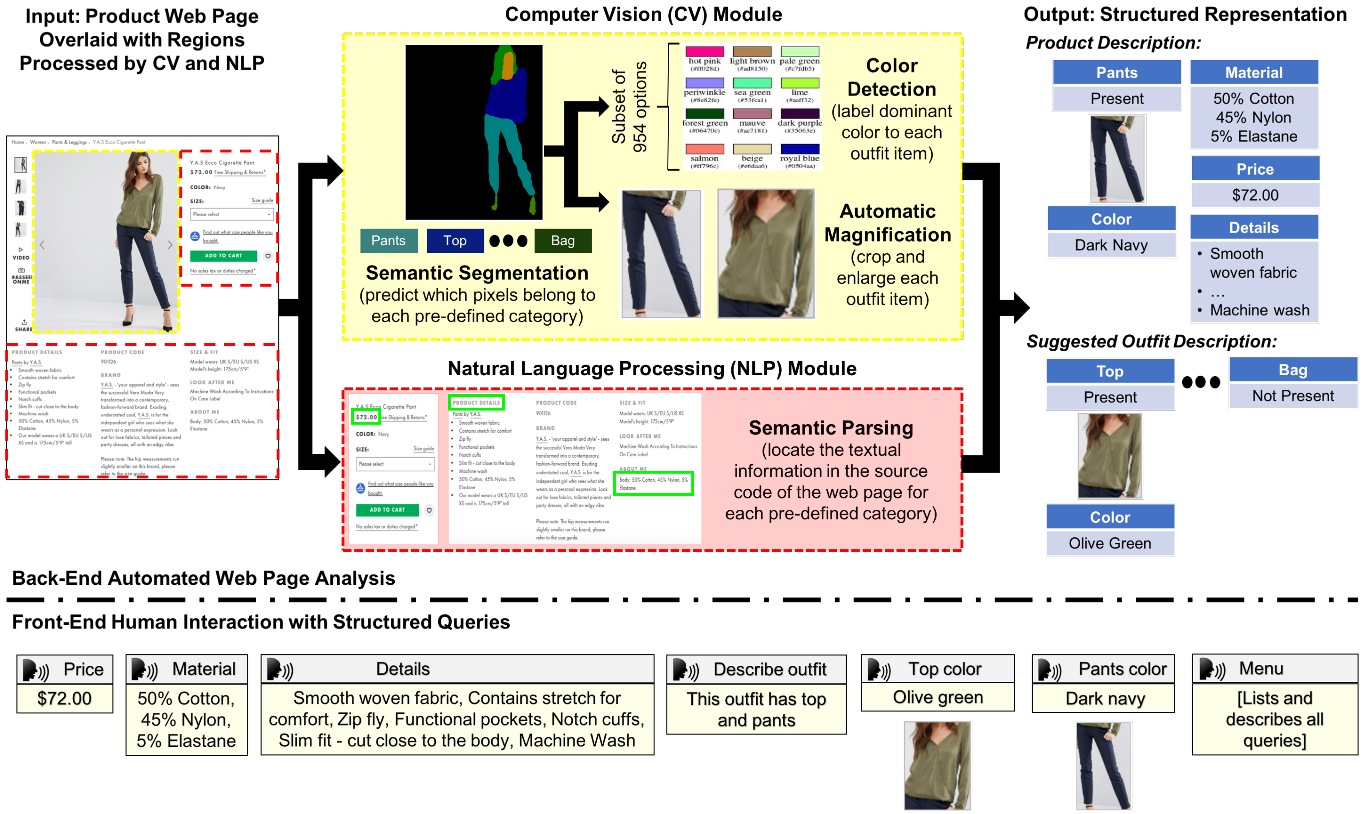


Figure 3. Overview of BrowseWithMe, which converts multimodal information (text and images) on a product web page into a structured representation that a user can interactively query to learn what (s)he wants to know about a product and the suggested outfit for the product.

cessibility challenges identified in the empirical investigation, we focus here on providing an automated assistant to help a user more easily obtain information about each product online. In other words, we propose an automated online assistant to help "browse" products, which we call BrowseWithMe.

BrowseWithMe enables a user to ask for and receive only the desired information from a web page. Figure 3 illustrates an overview of this system. On the back-end, BrowseWithMe takes as input a product web page and outputs a structured, hierarchical description about the product. It consists of a parallel set of modules that automatically re-organizes the multimodal content (e.g., text and images) on the web page into a consistent hierarchical structure and fills in missing information (i.e., image description). This approach complements existing web page segmentation approaches [49] by demonstrating how to automatically fill in missing information (e.g., image descriptions). On the front-end, BrowseWithMe, uses the structured representation as the basis for an intelligent online assistant that responds to user requests for information. We describe this system in greater detail below.

Back-end: Structured Representation

An important question to address when designing an interactive intelligent agent is "what type(s) of information do users want?" For our prototype, we identified four product-based features deemed important during the experimental investigation: price, material, details, and color. We also identified the following features to describe the outfit image on a product web page: report the article(s) of clothing present and the color of each article. These features reflect the broader desire of all participants to find outfits that match (i.e., Table 2; Q4 and Q5), and more specifically responds to Lily who shared "*I miss browsing through magazines to get outfit inspiration!*" The importance of these chosen features is reinforced by the findings of prior work [6, 13, 14, 20, 44, 46].

To generate the aforementioned hierarchical, structure that characterizes a product, we link a natural language processing (NLP) module and computer vision (CV) module. The NLP module is a simple parser that locates three factors about the product in the web page source code: price, material, and description. The CV module, described in greater detail below, automatically generates a description of the entire outfit shown in a product image. We implemented this back-end using a Python flask based web server. Automatically detecting such textual and visual information is the key idea that enables BrowseWithMe to relieve shoppers of the current challenge that they must learn different navigation behaviors for different shopping web sites to locate the pertinent information.

The key technical component that enables self-paced, hierarchical exploration of the different elements of an image is a semantic segmentation module. While existing automated assistive technologies [28, 35, 48] use image classification techniques that deliver a one-size-fits-all description for an image, we instead use a semantic segmentation module which automatically identifies the regions in an image belonging to the different clothing items (as exemplified in Figure 3). Our implementation supports automatically identifying ten categories of clothes. Included are five articles that commonly are in outfits: top, pants, dress, skirts, and shoes. Also included are five accessories: hat, sunglasses, belt, bag, and scarf. Sunglasses are of particular interest since people with vision loss often wear them both to safeguard what eyesight remains as well as to avoid the discomfort due to increased sensitivity to light. Hats also are commonly worn by those with sight loss to either avoid the discomfort from direct sunlight as well as to help detect overhead obstacles (e.g., tree branches). We implemented this module using CAFFE [26] to train a fully convolutional neural network on 9,000 annotated images [34].

The output semantic segmentation map is the foundation that BrowseWithMe then uses to assign a color to each detected item in the outfit, as exemplified in Figure 3. Specifically, each outfit item is assigned the color in the xkcd palette [4] that is closest in the Lab color space to the average pixel value from all pixels belonging to that item. We chose the xkcd palette because it uses widely-known color names. As an example of the human-like, descriptive richness of this palette, various shades of “red" may be identified as “lipstick red" or “brick red". Such nuanced detail about color can be especially valuable for individuals who lose sight later in life and so may be familiar with such distinctions.

The output semantic segmentation map also serves as the foundation to support smart magnification, a technique designed to improve the accessibility of magnification tools for low vision users. Specifically, for each detected item in an outfit, the system slightly enlarges the bounding box around all pixels that belong to it and crops out the remainder of the image, as exemplified in Figure 3. We intentionally restrict the content to only show the clothing of interest in order to avoid distracting users from the background or other articles of clothing.

Front-end: Automated Response to Queries

On the front-end, users can ask for and receive only the desired information using a single, consistent interface across different clothes shopping web sites. To demonstrate the general value of BrowseWithMe, we implement our prototype to support three popular web sites: (1) ASOS, which carries 850 brands of clothes, (2) H&M, a popular retailer in the United States and (3) Forever 21, a popular retailer in the United States. Customizing the system for each web site only requires updating the semantic parser to locate appropriate HTML identifiers.

BrowseWithMe responds to the following queries about the product: “price", “material", “details", and “color". It also responds to “describe outfit" by listing all items present in the product image (e.g., top, pants, hat). The user also can learn the color of any article of clothing and simultaneously have it magnified on the screen with the query “[article of clothing] color" (e.g., skirt color). BrowseWithMe magnifies an item to three times its size in the original image. The NLP module supplies the information for three of the queries (“price", “material", “details") and the CV module supplies the information for the remaining queries. Finally, BrowseWithMe responds to “menu" by listing all supported queries and describing the information provided by each query.

BrowseWithMe supports user interaction via typing and speaking. Typing may be preferred when users are in public spaces and do not wish to draw attention to themselves. Speaking, alternatively, offers hands-free convenience. Our front-end, client side of BrowseWithMe is implemented using Javascript. We used Chrome browser’s built-in voice recognition and dictation API to support voice commands.

USABILITY TESTING

We next investigate the value of BrowseWithMe in assisting visually impaired users to shop online.

Interview Design

We conducted a mixed method interview with the same eight individuals from the empirical investigation, a valuable precursor for assessing whether the system addressed the challenges and requests they communicated during the earlier interviews. Each user study lasted approximately one hour.

We began each interview with a moderated, problem-discovery usability test in which participants explored product web pages on three different web sites (i.e., ASOS [7], H&M, and Forever 21). We first explained to each user that BrowseWithMe is an interactive tool that enables users, when browsing clothes online, to give commands to learn about a product or the outfit associated with the product. We then provided a short training on how to interact with the system (i.e., price, material, details, describe outfit, [item] color, and menu queries) and advance the product web pages with the keyboard. As part of this training, we walked participants through requesting product information from two web pages, during which we reminded them of the queries they could use. Subsequently, we asked them to use the system independently to obtain information from six additional product web pages. For this test, we had the users engage with the system using voice commands.

We then asked each participant to answer a series of questions about their experiences with BrowseWithMe. Included were 12 open-ended questions and 12 Likert scale questions (1= Strongly Disagree; 2= Disagree; 3= Neutral; 4=Agree; 5=Strongly Agree). The Likert scale questions related to the Likert scale questions asked during the empirical investigation. To support data analysis, we collected field notes as well as audio and video recordings of each participant’s use of the system. Our emphasis in our analysis was to learn how each participant (1) completed requested tasks, (2) used voice queries to learn about products, and (3) might want to expand upon BrowseWithMe’s skills.

Impact Specific to Online Shopping Experience

We first discuss the impact of BrowseWithMe on making shopping web sites more accessible. Towards this aim, we report below both on how participants learned to interact with BrowseWithMe as well as their experiences with respect to the three aforementioned design recommendations we used to inform the design of our automated, online shopping assistant.

Learnability

All participants were able to complete the task of using BrowseWithMe to independently obtain information about six products, after completing training on two product web pages. Most participants remembered the commands after the two training rounds, although Annie recommended having a braille cheat sheet available and Jackie recommended having a way of accessing the command list in addition to the voice command "menu." Every participant was using the commands fluently by the final product web page.

Consistent Structure

The primary benefit the participants identified about the speech-driven interface was the ability to efficiently navigate to the information about the product that is important to them using the same five commands on any web site. For example, Jackie shared "*When I am product searching, I don't want all of the product information. It is nice to be able to ask for what I want to know.*" As noted by Bob, "*the most useful things are the details and the outfit description with access to colors*", reinforcing our broader observation that participants used the *details* command most often as their first action.

Complete Information

While the existing commands proved useful, most participants also requested more commands. They wanted to quickly learn about available color and size options for a product, brands, and the gender ascribed to the product. They also requested more fine-grained detail. For example, Bob requested "*more detailed color analysis with patterning as the algorithms get better; skirt length, skirt style (pencil, a-frame); sleeve length (1/4 length; 3/4 length); better description of neckline; dress – is it backless, U-cut; how does it fit around the waist – low waist; slit – does it have a slit which goes to the knee or upper thigh; does it have holes; all this would be good; descriptions for all graphics showing the item or outfit would be good – side, front, and back view.*" Julia requested that "*I could say, "are there pockets and, as if it [BrowseWithMe] had eyes, say there is a pocket." That would be amazing!*" Several participants also wanted to learn about the product material texture and blends as well as have an accessible sizing chart. Additionally, three participants requested the ability to interrupt the system, suggesting that the participants want to have real-time control in deciding when they have all the information they want. As noted by Bob, *"In its current form I can't talk over it... I need to wait for it to stop talking before I can give a new command."*

Hierarchical Structure

The four participants in our study with low vision appreciated the magnification feature to quickly zoom into any clothing item. Still, there were mixed reviews of it in its current form. Cathy said, *"I really liked when it made the image bigger. It could be made better by making the image in HDMI. I use a mouse a lot to scroll.* Sam indicated he would use the feature if it was there, but he might just default to using the magnification tool on his computer or phone. Vivian wanted the image to be magnified even more for it to be of use to her.

More generally, supporting a broader range of commands to support online clothes shopping necessitates a hierarchical structure to support a user to be efficient when shopping.

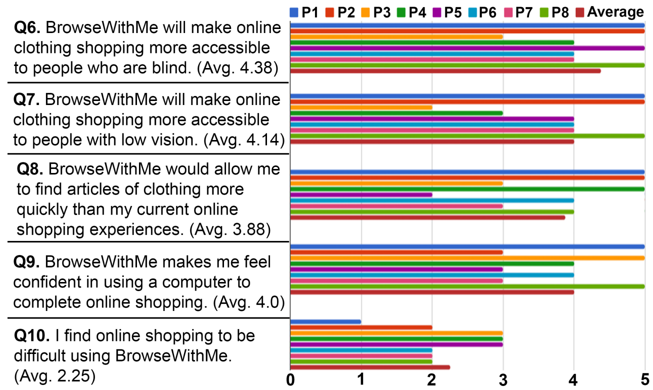


Figure 4. Study participants answers to several Likert scale questions asked about their experiences with BrowseWithMe; 1=Strongly Disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly Agree.

Impact on General Shopping Experience

We next examine the impact of BrowseWithMe on the participants' shopping experiences, reporting findings with respect to the five values summarized in the empirical investigation. It is noteworthy that we intentionally excluded addressing one value, "Social Connection", from our system, as we initially believed it would require a huge technical effort of building an online social network. As discussed below about this value, and other values, participants were very generous in providing valuable feedback for how to further improve BrowseWithMe to support their values and enhance their experiences.

Efficiency and Ease: Most participants thought BrowseWithMe would allow them to find clothes more quickly online (i.e., Q8; Figure 4). Most participants also agreed or strongly agreed that online shopping will be more accessible with BrowseWithMe (i.e., Q6 and Q7; Figure 4). No participant reported that BrowseWithMe was difficult to use (i.e., Q10; Figure 4). Anecdotally, Sam shared "*I think for myself, reading a lot of text can be tiresome for eyes, so leaving those and moving to audio can be helpful, and knowing the details of products and specs can be useful...I like having the option of voice.*" He also suggested a further benefit of speech-driven interaction is the ability to multitask: *"This could be really useful if I wanted to use voice commands to search and shop for different items and do something else at the same time."*

Information Access and Trust: Lily and Jackie, who recently lost their sight, related the experience of using BrowseWithMe to catalogue shopping. In Jackie's words, *"it is more like looking at an ad than at a specific product; it reminded me of reading the Sunday paper...more like browsing to put together a complete outfit. I didn't expect that."* Vivian shared the system allowed her to learn what fashions are current, while removing the burden of sifting through too much information.

Independence: Most participants reported that BrowseWithMe makes them feel confident in using a computer to shop online (Q9; Figure 4).

Social Connection: While our current system was not designed with the intention to support social connection online, we heard from Sam, Jackie, Cathy, and Vickie a nice sugges-

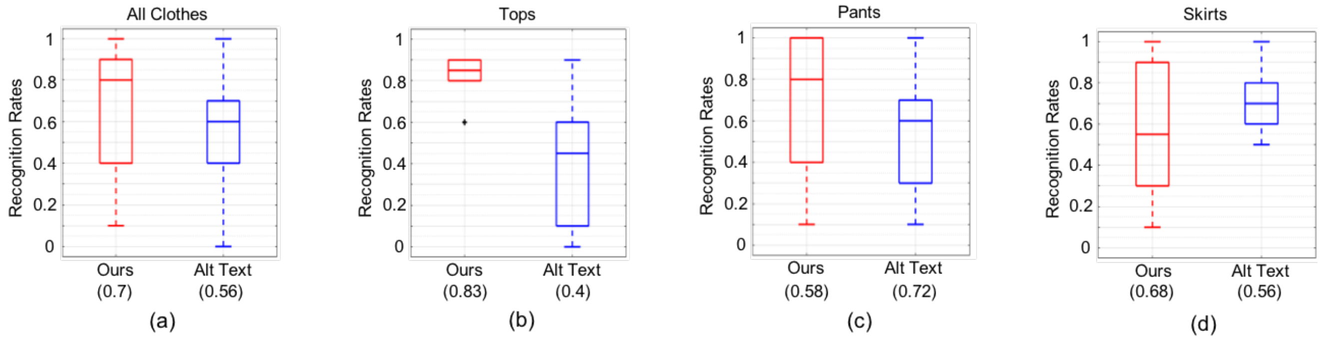


Figure 5. Human voting accuracy in choosing the correct image from five options, when given an image description from BrowseWithMe and Alt text. Shown are results for (a) 30 voting tasks which consist of (b-d) 10 voting tasks per clothing item for "tops", "pants", and "skirts" respectively. Each score represents the fraction of ten people who chose the correct image from five options. The central marks of the boxes denote the median values, box edges denote the 25th and 75th percentiles values, whiskers denote the adjacent value to the data point that is greater than one and a half times the size of the inter-quartile range, and black cross-hairs denote outliers. Also shown below the plots are the mean values. Alt text leads to accurate recognition only slightly more than half the time. BrowseWithMe yields a great improvement over Alt text; e.g., 20 percentage point increase in the median score.

tion for how to introduce this capability. Specifically, they requested a feature to support sending pictures of clothing to friends for feedback. Such a feature would mimic our participants' experience while shopping in brick-and-mortar stores of engaging with others to get feedback, good deals, etc.

Freedom of Expression and Personalization: As one participant noted, "*it [BrowseWithMe] could help me figure out what I want, what is out there... Sometimes I don't know what I want or what styles are new.*" Julia suggested a further potential of the system; *"If you could link to each article an outfit, you could satisfy the retail therapy of browsing. This system enables you to not have to know what you are looking for."* At the end of her interview, Jackie stated "*I think it [BrowseWithMe] would help me put together an outfit and stylize...as opposed to just getting a shirt or jeans.*"

TECHNICAL EVALUATION: IMAGE DESCRIPTIONS

We next evaluate the ability of BrowseWithMe to automatically generate an accurate description of a product image. This addresses participants' frustration that Alt text often is missing for images, leaving them to a tedious process of parsing the content on the rest of a web site to try to learn whether information (such as product color) is available elsewhere. Of note, prior work that explored the use of automatically generated Alt text [35] concluded that blind people should mistrust resulting image descriptions. Our work highlights a different finding in the limited domain of outfit images on shopping web sites.

Method

Our experiment is designed to evaluate and compare image descriptions that come from BrowseWithMe and Alt text.

Task. Study participants were shown a text description and asked to identify the best matching image from five image options. The five image options consisted of one true match and four incorrect matches. A study participant voted by selecting a radio button immediately below an image.

Participants. We recruited sighted participants, since our aim is to assess if a description accurately matches an image. We employed crowd workers on Amazon Mechanical Turk who previously completed at least 500 tasks and received a 95% approval rating. We only accepted crowd workers from the US to minimize concerns about language barriers.

Dataset. We created a dataset to support evaluating image descriptions. We chose true image matches by collecting 30 images from two shopping web sites (i.e., ASOS and Selfridges). We used three search queries—"skirt", "pants", and "top"—to support evaluation for a variety of products, and collected 10 images per search query. For each of the resulting 30 images, we collected a text description from BrowseWithMe (by concatenating each article of clothing with its color, separated by commas; e.g., "black top, almost black bag, light peach skirt") and from Alt text (when no Alt text was available, we used the title listed on the web site; e.g., "Leather Look Stretch Skinny Pants"). Next, we identified incorrect image matches by collecting an additional four images from ASOS and Selfridges that we paired with each true image. We employed the same search queries used to collect the true image—"skirt", "pants", and "top". Consequently, incorrect image options typically resembled the true option by including the same elements of an outfit (e.g., all five images show a skirt and top) and occasionally showed similar colored outfits.

In total, we collected 150 images to support 60 voting tasks. The dataset supported 30 voting tasks on BrowseWithMe descriptions and 30 voting tasks on Alt text descriptions.

Experimental Design. To minimize concerns about fatigue affecting a worker's performance, we only included 20 voting tasks per job. Included were 10 voting tasks with the BrowseWithMe descriptions and the corresponding 10 voting tasks with the Alt text for the same images. In order to capture possible disagreements about which is the best matching image for a given description, we collected votes from 10 unique workers per description. To avoid possible order biases, for each worker, we set a new randomized order of the 20 voting tasks per job and order of the five images per task.

Evaluation. For each image description, we computed the fraction of 10 crowd workers who chose the correct image. Thus, our quality measure for each image description is represented by a score ranging from 0 to 1 with higher scores indicating higher quality. To assess the significance of differences in

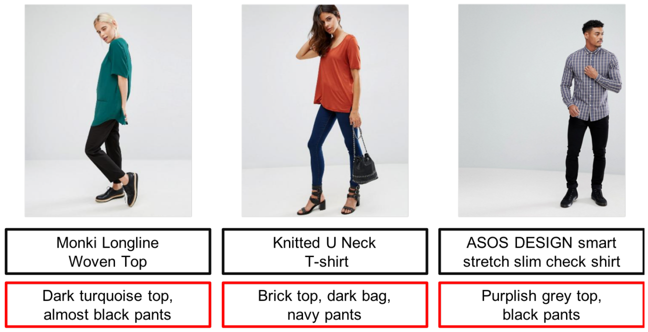


Figure 6. Comparison of image descriptions from BrowseWithMe (in red rectangles) and Alt text (in black rectangles).

quality for BrowseWithMe descriptions and Alt text, we used the bootstrap test [41] to compute the probability of obtaining the two sets of observed scores by chance.

Results

Overall, crowd workers recognized images better using automatically-generated descriptions from BrowseWithMe than using Alt text (**Figure 5a**); i.e., the median improves by 20 percentage points from 60% to 80%. The chance of obtaining the two sets of observed results by chance is low (p = 0.06). **Figure 6** exemplifies the advantage of BrowseWithMe as well as its limitation with handling non-solid colored items.

We also observed the CV module performs well in absolute terms; i.e., 80% median score for all voting tasks (**Figure 5a**). We attribute this success in part to images on shopping web sites often having clean, solid colored backgrounds with similarly structured content in the foreground. Specifically, we hypothesize that the restricted scope of image content in the online shopping domain makes the prediction task easier for CV algorithms. Our findings offer promising evidence that online shopping is a domain where artificial intelligence algorithms can be trusted and so of great value.

For specific products, we found BrowseWithMe performs best for tops (**Figure 5b**), followed by pants (**Figure 5c**), and finally skirts (**Figure 5d**). Many mistakes arise due to the segmentation algorithm confusing skirts for dresses and inaccurate color assignments (e.g., labeling a white top as pale grey). Our findings highlight a need to further improve CV algorithms for detecting outfit items and characterizing colors.

DISCUSSION

While the accessibility of online shopping depends on many factors, our findings suggest the browsing capability is an important factor. Below we discuss areas for future work.

Broadening Scope of BrowseWithMe

All participants expressed an interest in extending BrowseWithMe for use throughout their entire shopping experience, from the moment they start to search for an item to checkout. For example, Annie said it would be helpful to connect this system to checkout so she could quickly review the items in her shopping cart. Vivian asked for a *Buy This Item!* command.

Our user studies also uncovered an interest from most participants to support online shopping for products besides clothing. Most participants we interviewed use Amazon to shop online for items such as food, household supplies, and electronic products. While these product categories may not carry the same type of personal significance as clothes shopping, they pose similar technical challenges in that shopping web sites typically show multimodal web pages (i.e., product images with associated text). Generalizing BrowseWithMe to other shopping web sites and domains requires engineering the backend system; e.g., swap in a different semantic segmentation module and semantic language parser. We plan to investigate deploying BrowseWithMe for different products to evaluate and compare users' experiences in different shopping domains.

Broadening Scope of User Studies

While our user studies offer insight from a breadth of users across a diverse population, a valuable next step will be to pursue a deep examination of the interests and experiences for different segments of the population; e.g., based on visual impairment—born blind, went blind later in life, versus low vision; age; and technical expertise. We suspect that the kind of interests and perceived value of online shopping technology will vary across different user groups. For example, we heard in our user studies that more experienced screen reader users preferred BrowseWithMe to speak faster whereas less experienced screen reader users preferred BrowseWithMe to speak slower. In addition, while the magnification tool is only of use to people with low vision, it could be improved by designing different magnification schemes for objects of different size (e.g., sunglasses versus a dress) as well as different screen sizes (e.g., mobile phone versus computer screen).

Thus far, we studied a web browsing experience with known users to design and identify design improvements for BrowseWithMe. Studies with a large number of anonymous users in a real-life, online shopping context will be important to further generalize BrowseWithMe for use at scale “in the wild".

CONCLUSIONS

We explored the problem of designing technology to assist people with visual impairments to shop for clothes online as a non-trivial example for understanding and improving the accessibility of online shopping at large. In-person interviews motivated the practical value of this problem by illustrating the importance and current inaccessibility of clothes shopping. In response, we proposed BrowseWithMe, which departed from treating people as passive listeners of unparsed information to active solicitors of the specific information they are seeking. BrowseWithMe automatically converts multimodal input data into a structured representation in order to overcome the inconsistent shopping experiences provided by different web sites. Findings from our Usability Testing and Technical Evaluation validated that BrowseWithMe could improve users' online shopping experience and fill in missing information. We publicly share our code at `https://github.com/kothariesha/BrowseWithMe`.

Acknowledgments. We thank study participants and crowd workers for their valuable time. This work is supported by NSF awards to D. G. (IIS-1755593) and K. G. (IIS-1514118) as well as an Amazon Research Award to K. G.

REFERENCES

1. Be My Eyes. http://www.bemyeyes.org/.

2. http://webaim.org/projects/screenreadersurvey/.

3. http://www.afb.org/afbpress/pubnew.asp?DocID=aw170208.

4. xkcd Color Palette. https://xkcd.com/color/rgb/.

5. L. Von Ahn and L. Dabbish. 2004. Labeling Images With a Computer Game. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 319–326.

6. R. M. Alali. 2017. Case study of visually impaired individuals' preferences of the availability of braille clothing labels in shopping and selection of apparel. (2017).

7. ASOS. Online shopping for the Latest Clothes & Fashion. `http://us.asos.com/`. (Accessed on 09/10/2017).

8. M. Avila, K. Wolf, A. Brock, and N. Henze. 2016. Remote Assistance for Blind Users in Daily Life: A Survey about Be My Eyes. In *International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*. ACM, 85.

9. J. Bigham. 2014. Making the Web Easier to See with Opportunistic Accessibility Improvement. In *ACM symposium on User Interface Software and technology (UIST)*. 117–122.

10. J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *ACM symposium on User interface software and technology (UIST)*. 333–342.

11. J. P. Bigham, R. S. Kaminsky, R. E. Ladner, O. M. Danielsson, and Gordon L. Hempton. 2006. Webinsight: Making Web Images Accessible. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

12. J. P. Bigham, I. Lin, and S. Savage. 2017. The Effects of "Not Knowing What You Don't Know" on Web Accessibility for Blind Web Users. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

13. E. Brady, M. R. Morris, Y. Zhong, S. White, and J. P. Bigham. 2013. Visual Challenges in the Everyday Lives of Blind People. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 2117–2126.

14. M. A. Burton, E. Brady, R. Brewer, C. Neylan, J. P. Bigham, and A. Hurst. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. In *ACM SIGACCESS conference on Computers and accessibility (ASSETS)*. 135–142.

15. X. Chen, H. Fang, T. Y. Lin, R. Vedantam, S. Gupta, P. Dollár, and C. L. Zitnick. Microsoft COCO captions: Data collection and evaluation server. arXiv preprint arXiv:1504.00325 (2015).

16. T. B. Dinesh, S Uskudarli, S. Sastry, D. Aggarwal, and V. Choppella. 2012. Alipi: A framework for re-narrating web pages. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*.

17. J. Fraser and C. Gutwin. 2000. A framework of assistive pointers for low vision users. In *ACM SIGACCESS conference on Computers and accessibility (ASSETS)*. 9–16.

18. Batya Friedman, Peter H Kahn, Alan Borning, and Alina Huldtgren. 2013. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*. Springer, 55–95.

19. C. P. Gharpure and V. A. Kulyukin. 2008. Robot-assisted shopping for the blind: issues in spatial cognition and product selection. *Intelligent Service Robotics* 1, 3 (2008), 237–251.

20. D. Gurari, Q. Li, A. J. Stangl, A. Guo, C. Lin, K. Grauman, J. Luo, and J. P. Bigham. 2018. VizWiz Grand Challenge: Answering Visual Questions from Blind People. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3608–3617.

21. S. Harper and A. Q. Chen. 2012. Web accessibility guidelines. *Web accessibility guidelines* 15, 1 (2012), 61–88.

22. C. Harrison and A. K. Dey. 2008. Lean and zoom: proximity-aware user interface and content magnification. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 507–510.

23. T. Hironobu, S. Kawanaka, M. Kobayashi, T. Itoh, and C. Asakawa. 2008. Social Accessibility: Achieving Accessibility through Collaborative Metadata Authoring. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. 193–200.

24. M. Jas and D. Parikh. 2015. Image Specificity. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2727–2736.

25. H. Jethava, S. Zafar, and M. Saini. 2014. Electronic Shopping Cart Facility for Blind People Using USB Firmware. In *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4. 647–651.

26. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*.

27. Y. Jing, D. Liu, D. Kislyuk, A. Zhai, J. Xu, J. Donahue, and S. Tavel. 2015. Visual Search at Pinterest. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1889–1898.

28. H. Kacorri, K. M. Kitani, J. P. Bigham, and C. Asakawa. 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In *ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, 5839–5849.

29. S. Kesh, S. V. Venkateshwar, and A. U. 2017. Shopping By Blind People: Detection of Interactions in Ambient Assisted Living Environments using RFID. In *International Journal of Wireless Communications and Networking Technologies*, Vol. 6. 7–11.

30. D. Kislyuk, Y. Liu, D. Liu, E. Tzeng, and Y. Jing. 2015. Human Curation and Convnets: Powering Item-to-Item Recommendations on Pinterest. arXiv preprint arXiv:1511.04003. (2015).

31. R.L. Kline and E.P. Glinert. 1995. Improving GUI accessibility for people with low vision. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 114–121.

32. A. Kovashka, D. Parikh, and K. Grauman. 2012. Whittlesearch: Image search with relative attribute feedback. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2973–2980.

33. W. S. Lasecki, R. Kushalnagar, and J. P. Bigham. 2014. Legion Scribe: Real-time Captioning by Non-experts. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. ACM, 303–304.

34. Xiaodan Liang, Chunyan Xu, Xiaohui Shen, Jianchao Yang, Si Liu, Jinhui Tang, Liang Lin, and Shuicheng Yan. 2015. Human Parsing With Contextualized Convolutional Neural Network. In *The IEEE International Conference on Computer Vision (ICCV)*.

35. H. MacLeod, C. L. Bennett, M. R. Morris, and E. Cutrell. 2017. Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. In *ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, 5988–5999.

36. N. Paisios, L. Subramanian, and A. Rubinsteyn. 2012. Choosing which clothes to wear confidently: A tool for pattern matching. In *Workshop on Frontiers in Accessibility for Pervasive Computing. ACM*.

37. H. Petrie and O. Kheir. 2007. The Relationship between Accessibility and Usability of Websites. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 397–406.

38. C. Power, A. P. Freire, H. Petrie, and D. Swallow. 2012. Guidelines are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 433–442.

39. G. VRJ Prasad, S. Chimalakonda, V. Choppella, and Y R. Reddy. 2017. An Aspect Oriented Approach for Renarrating Web Content. In *ACM Innovations in Software Engineering Conference*. 56–65.

40. E. Salisbury, E. Kamar, and M. Morris Ringel. 2017. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. (2017).

41. M. D. Smucker, J. Allan, and B. Carterette. 2007. A Comparison of Statistical Significance Tests for Information Retrieval Evaluation. In *ACM conference on Conference on Information and Knowledge Management*. 623–632.

42. L. Stearns, V. DeSouza, J. Yin, L. Findlater, and J. E. Froehlich. 2017. Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

43. Anselm Strauss and Juliet M Corbin. 1990. *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Publications, Inc.

44. S. Szpiro, Y. Zhao, and S. Azenkot. 2016. Finding a store, searching for a product: a study of daily challenges of low vision people. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 61–72.

45. L. von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum. 2006. Improving Accessibility of the Web with a Computer Game. In *SIGCHI Conference on Human Factors in computing systems (CHI)*. 79–82.

46. M. A Williams, A. Hurst, and S. K. Kane. 2013. Pray Before You Step Out: Describing Personal and Situational Blind Navigation Behaviors. In *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

47. T. Winlock, E. Christiansen, and S. Belongie. 2010. Toward real-time grocery detection for the visually impaired. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 49–56.

48. S. Wu, J. Wieland, O. Farivar, and J. Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In *CSCW*. 1180–1992.

49. Y. Yesilada. 2011. Web page segmentation: A review. *Northern Cyprus Campus* (2011).

50. S. Yuan, Y. Tian, and A. Arditi. 2011. Clothing matching for visually impaired persons. *Technology and disability* 23, 2 (2011), 75–85.

51. A. Zhai, D. Kislyuk, Y. Jing, M. Feng, E. Tzeng, J. Donahue, Y. L. Du, and T. Darrell. 2017. Visual Discovery at Pinterest. arXiv preprint arXiv:1702.04680. (2017).

52. Y. Zhao, S. Szpiro, J. Knighten, and S. Azenkot. 2016. CueSee: Exploring Visual Cues for People with Low Vision to Facilitate a Visual Search Task. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 73–84.

53. Y. Zhong, W. S. Lasecki, E. Brady, and J. P. Bigham. 2015. Regionspeak: Quick comprehensive spatial descriptions of complex images for blind users. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 2353–2362.