Designing an Eyes-Reduced Document Skimming App for Situational Impairments

Taslim Arefin Khan
University of British Columbia
Vancouver, Canada
takhan@cs.ubc.ca

Dongwook Yoon
University of British Columbia
Vancouver, Canada
yoon@cs.ubc.ca

Joanna McGrenere
University of British Columbia
Vancouver, Canada
joanna@cs.ubc.ca

ABSTRACT
Listening to text using read-aloud applications is a popular way for people to consume content when their visual attention is situationally impaired (e.g., commuting, walking, tired eyes). However, due to the linear nature of audio, such apps do not support skimming—a non-linear, rapid form of reading—essential for quickly grasping the gist and organization of difficult texts, like academic or professional documents. To support auditory skimming for situational impairments, we (1) identified the user needs and challenges in auditory skimming through a formative study (N=20), (2) derived the concept of "eyes-reduced" skimming that blends auditory and visual modes of reading, inspired by how our participants mixed visual and non-visual interactions, (3) generated a set of design guidelines for eyes-reduced skimming, and (4) designed and evaluated a novel audio skimming app that embodies the guidelines. Our in-situ preliminary observation study (N=6) suggested that participants were positive about our design and were able to auditorily skim documents. We discuss design implications for eyes-reduced reading, read-aloud apps, and text-to-speech engines.

Author Keywords
Situational impairments; skim reading; eyes-free; eyes-reduced; text-to-speech; mobile device; accessibility; design guidelines; interactive prototype.

CCS Concepts
•**Human-centered computing** → **Interaction design process and methods;** *Accessibility systems and tools;* **Gestural input; Auditory feedback;**

INTRODUCTION
The consumption of audio media is growing as digital streaming services and mobile devices become pervasive. This trend started with the popularity of audio content produced and curated specifically for listening (e.g., podcasts, audiobooks). Over the past several decades, text-to-speech (TTS) technology has improved dramatically. TTS-based *read-aloud apps* represent a new trend in auditory media consumption. Such apps can narrate news articles, e-books, and webpages for users to listen to in situations where visual reading is challenging or impossible (e.g., on-the-go, tired eyes). However, listening to audio content is typically a linear and slow experience [42, 50]. To date, such apps lack support for a dynamic and rapid style of reading called *skimming*. Our work strives to explore, design for, and support *auditory skimming for situational impairments*.

ACM ISBN 978-1-4503-6708-0/20/04.
http://dx.doi.org/10.1145/3313831.3376641

Skimming is a rapid, selective, and non-sequential form of reading [19, 53]. People skim a document when they want to quickly get the gist of it, look for specific information in it, and learn its structural organisation. Quickly skimming over a document is a time-efficient way to assess whether the given content warrants the effort of a full read [1]. In today's media landscape where the amount of online content increases rapidly each year [39], skimming has become the new norm [62]. To keep up with even a fraction of what's available, people are using skim-reading more than ever before. Academics, just as one example class of reader, have noted the challenge of keeping up with the literature, given the explosion of publications [32].

Skim-reading, despite its importance, is limited in terms of how, when, and where it can be done. More specifically, there is no straightforward way for users to skim when they are situationally impaired, unable to look at the screen. Skimming depends heavily on a reader's *visual* interactions with texts: rapid eye movement wandering over the text [9], glancing to spot keywords, selectively reading sentences relevant to the goal, and jumping around the text until the information need is met. A prime example is that it is not possible to skim read in "on-the-go" situations using existing read-aloud apps, in the way that one can linearly listen to podcasts or even audiobooks. There is an open design problem of translating the visual interactions in skim-reading into a mode of interactions that depend less on visual attention.

We address this research problem by selectively combining auditory and visual reading modes. We call this approach *eyes-reduced* design to juxtapose it with eyes-free [63]. Eyes-reduced skimming enables a given document to be skimmable primarily auditorily and eyes-free but with the option to be glanced down at, to see visual content, such as embedded images, on an as-needed basis. Skimming auditorily should help individuals who are situationally impaired, using their visual attention for other tasks, such as walking, or avoiding

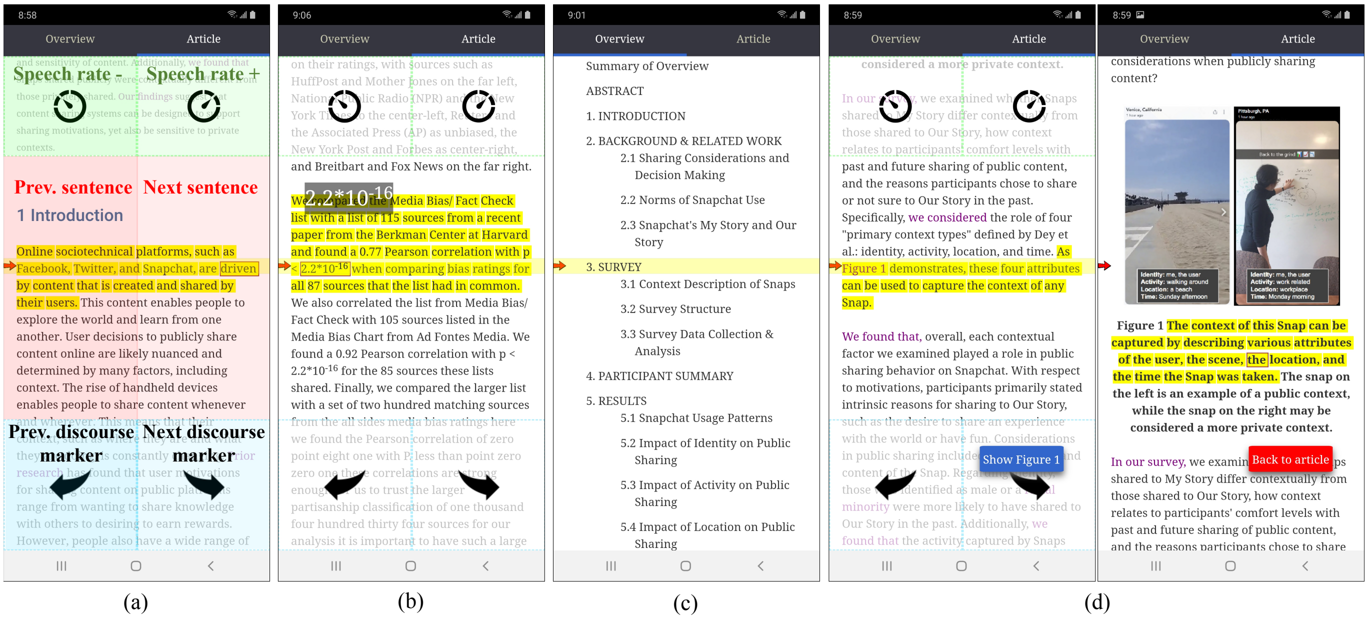

Figure 1. Skimmer interfaces; (a) Tapping on each of three gesture regions triggers speech rate control (top), sentence navigation (middle), and jumping between discourse markers respectively (bottom); (b) When encountering a number, Skimmer narrates it in an easy-to-understand rounded format (see the section on the Design of Skimmer for details) and prompts the enlarged numeric text on the screen; (c) The Overview page outlines the structural organization; (d) When a Figure/Table is referred to in the text, Skimmer gently nudges the user with haptic feedback and gives them an opt-in option to jump to the referent (left). Upon opt-in, Skimmer shows the Figure/Table and narrates the caption (right).

motion sickness from reading in a moving vehicle, such as a bus or train. The commuting context is particularly motivating, as the increasing commute time is a global phenomenon [27].

To address our research problem, we: (1) conducted a user study (N=20) in a simulated on-the-go context to understand the challenges and design considerations for auditory skimming, (2) developed a set of design guidelines for eyes-reduced skimming, (3) designed and implemented Skimmer, a read-aloud app for eyes-reduced skimming (Figure 1), and (4) conducted a preliminary evaluation (N = 6) of Skimmer.

Our work contributes: (1) the concept of eyes-reduced skimming for situational impairments, (2) the design guidelines for eyes-reduced skimming, and (3) the design of Skimmer, a system that implements those guidelines.

RELATED WORK

This work is informed by previous studies on situational impairment, eyes-free and non-visual interaction systems, human listening capabilities, audio interfaces, and empirical works on skimming behaviors.

Situational Impairment and Eyes-free Interaction

Situational Impairment (SI) is the temporary deterioration of human abilities caused by situational factors [47]. A common example of SI is the challenge people face when using a mobile device on the go (e.g., walking or driving) [60]. Well-documented situational factors that can induce SI include the encumbrance of carrying an object, cold temperature, ambient noise, being mobile, mood and stress level [41, 45]. People often choose to use their mobile device eyes-free, without lowering their head to the screen, due to varying environmental, social, device specific, and personal factors (e.g., extreme lighting conditions, being in a meeting, tiny screen, self-enthusiasm). To address these challenges, researchers have proposed several approaches, such as improving touch accuracy by generating user-specific touch models [40], enlarging targets [28], CrashAlert [25] for eyes-busy mobile interaction while walking, and eyes-free unimanual bezel-initiated gestures [10].

A different line of research has sought to identify and address problems of reading in SI. Empirical studies have shown that reading in SI can result in lower comprehension and higher cognitive load [46, 55]. Vadas et al. suggested synthesized speech of texts as a viable solution that can yield a reading performance comparable to that of head-down visual reading [55]. SeeReader [14] extended this approach by supporting region-based (figures, tables, and paragraphs) navigation, reading aloud summary of regions, and notifying presence of images. Read4Me [64] is a prototype browser that leverages smartphone sensors to detect context switches. It automatically onboards the user to a hands-free mode when the user starts moving. Our work extends beyond existing studies of reading on the go to enable skim-reading, a type of reading that has been unsupported for SI.

Researchers have explored the potential of using audio and haptics in eyes-free interfaces. Vazquez-Alvarez and Brewster [56] explored eyes-free mobile multi-tasking, showing the effectiveness of spatial audio in divided- and selective-attention tasks. Other researchers designed audio-based eyes-free menu-selection interfaces [34, 66] or sonical enhancement of the existing menu interfaces such as radial menu or hand gesture [12]. Other studies have leveraged haptic feedback for eyes-free mobile device interaction and menu selection [7, 43]. PocketMenu [44] is a non-visual interaction technique with touchscreen menu items for on-the-go contexts, combining both audio (speech) and haptic feedback. In our work, we leverage both earcons and haptic feedback as multimodal navigation cues in an eyes-reduced skimming context.

Human Listening and Audio Interfaces

Researchers have explored the effects of TTS on both sighted and visually impaired (VI) populations. Bragg et al. [11] performed a study on the intelligibility of TTS with varying speech rates by sighted and VI participants. The study confirmed that VI people can understand synthetic speech at a higher rate than sighted people, and that sighted people can understand speech at a rate higher (297 Words Per Minute (WPM)) than normal speaking rates (120 - 180 WPM). Cohen et al. [17] recommended guidelines for Voice User Interface (VUI) design based on the nuance of human listening perceptions (e.g., breaking down a long list of items into groups of 3 or 4). Some of these guidelines may be transferable to document skimming in a SI scenario. We extend these works by putting forward novel findings from our empirical study (e.g., pausing narration when the user is navigating.)

In order to understand state-of-the-art read-aloud apps, we informally surveyed 13 off-the-shelf applications designed for both VI and sighted users. Examples include iBooks (based on Apple VoiceOver), TalkBack, NaturalReader, VoiceDreamReader, and @Voice Aloud Reader. Applications differ by types of navigation (spatial, semantic, temporal) and, within those types, granularity (page, chapter, sentence, 10 seconds, 30 seconds, and so on). Applications also differ in user interface designs—basic control support and spatial organization of menus and sub-menus. We used our survey to understand the feature sets of the existing read-aloud apps, to examine how well eyes-reduced skimming would be supported by each app, and to determine which app to use as the reference system in our Needs-Finding Study and Preliminary Evaluation.

Studies on Skimming

In other bodies of related work, researchers have explored skimming for two broad categories of materials: document and audio.

Document skimming

Masson [38] defined skimming as a "technique that is commonly associated with reading goals that involve the comprehension of only a subset of a story's content". The reader focuses on the information relevant to the goal of skimming. According to Adler and Van Doren [1], the main aim of skimming is to discover whether the material "requires a more careful reading." Studies such as [18, 19] found that skim-reading happens in patches, and a paragraph is the most plausible patch. When skimming, the reader first sticks to reading a paragraph but quickly moves to another paragraph when the rate of information gain from the current paragraph drops below a threshold. According to Marshall and Bly [9], skimming, scanning, and glancing are common reading strategies among people reading longer documents; they often embody *lightweight navigation* patterns such as looking ahead, looking back to re-read for context, and narrowing or broadening focus to an area. Ever since the widespread adoption of mobile devices and tablets, researchers have explored reading behaviors in digital environments, and found evidence of skimming as a popular strategy in digital reading [20, 24, 37, 53]. Our study benefits from the empirical understanding of skimming from the aforementioned studies, confirms their findings through empirical findings of our own, and contributes to further understanding of people's skimming strategies when experiencing SIs.

Researchers have proposed novel interfaces to aid document skimming. Structure-Aware Touch-Based Scrolling (SATS) [29] assists the reader to perform non-linear navigation. In SATS, the tablet screen is vertically divided into four parts, each corresponding to a different navigational unit, such as chapter, section, sub-section, and page. Spotlight [33] is an attention-optimised skim-reading tool that selects salient objects (e.g., title, headings, a figure) and displays them as a transparent overlay on the screen as the user scrolls. In our work, we take advantage of gesture-based navigation to support eyes-reduced skimming. In addition to Spotlight's findings, we found that discourse markers are also salient objects, as detailed in our Needs-Finding Study.

Audio skimming

SpeechSkimmer [6] supports auditory skimming of recorded speech by condensing pauses and structuring audio based on pause duration and non-speech sounds. Transcripts can be used as proxy to audio content for visual skimming and help spotting keywords by highlighting words with high confidence scores from speech recognizer output [57, 59]. Recent studies have focused on non-visual skimming for screen reader users. A series of studies by Ahmed et al. [2, 3] supported skimming of online content by automatically summarizing texts at multiple levels of abstraction. The authors extended their study to touch screen devices [4], and designed a set of novel gestures on top of VoiceOver's default gestures, such as pinch-in and out to control the level of summary. Machulla et al. proposed design implications for non-visual document skimming for VI individuals [35]. They proposed supporting navigation between macro-structures of the text (e.g., section title, paragraph, figure), changing pitch or adding auditory stimuli to emphasize micro-structures (e.g., highlighted text, bold or italicized words), and supporting dynamic speech rates. In our work, we show that some of these design ideas for the VI context are transferable to the SI context.

A NEEDS-FINDING STUDY FOR AUDITORY SKIMMING

We conducted a formative study to understand the needs of auditory skimming in a situationally impaired context. In particular, we sought to understand how might people skim auditorily, what are the challenges, what is working and not working, given the state-of-the-art read-aloud applications.

Method

We conducted the study in a controlled lab setting chosen to simulate a SI. We used a representative read-aloud app as a probe. Our study used a 2 (document formats) $\times$ 2 (article types) within-subjects design. In total, participants completed the task 4 times. The order of the conditions was fully counterbalanced.

Participants

We recruited 20 university graduate students (10 M/F) through convenience sampling. Participants had various backgrounds (e.g., Computer Science, Kinesiology, Occupational Therapy, Business). We recruited only those who reported familiarity with reading research articles, as a disparity between participants' familiarity with the reading tasks could introduce bias.

Task

We asked the participants to imagine that they are enrolled in a graduate level course and have a reading assignment for a class later that day in which they have to participate in a peer discussion session. In order to contribute to the discussion, they should skim the given reading materials under a time constraint while riding a bus to school. To reinforce the simulated SI, we asked participants to minimize visual attention to the screen, pretending that otherwise they might get nauseated. Looking at the screen was permitted to issue a command but they were instructed to look away when listening as best possible.

We asked participants to align their task objectives to the goal of skim-reading known from the literature: understanding the gist, topic/problem statements, key takeaways, and structural organisation of the article. Right after they finished the reading, we asked them four short comprehension questions about the above-mentioned objectives. Our intention of asking these questions was to put them in the shoes of a skim-reader situationally impaired rather than to get the correct answers to the questions.

Conditions

Document formats: Based on our survey on the existing read-aloud applications, we found that most applications support either one or both of the following layouts: (1) plain text layout that displays only text and discards any formatting, and (2) original layout that displays visual elements in the original arrangements of the PDFs. We studied both to understand if one is more favorable to skim-reading on the go.

Article types: We selected published articles of two types, academic papers and professional reports, as experimental materials for this study. We made sure that the chosen articles were isomorphic in terms of their topics, structures, and difficulty levels [30]. For research papers, we selected two from the Information, Communication & Society journal: A1 [13] and A2 [52] (# words: 8202 and 7260, respectively; difficulty level for both: college graduate). For professional reports, we selected two survey articles published by Pew Research Center: A3 [5] and A4 [51] (# words: 4664 and 4559, respectively; difficulty level for both: college). Overall, we followed rationales for selecting reading materials suggested in [18].

We imported PDF files of the articles into the reading app. All articles were formatted in the single column typeset to maintain consistency. Before that, we fixed the reading order of document elements using Adobe Acrobat's accessibility tool.

Figure 2. The experimental setup of our Needs-Finding Study. Participants were asked to stand in front of a screen playing ambient videos that simulate what commuters would see in a moving bus.

Apparatus

We used the VoiceDreamReader (VDR) app running on an iPhone 5S. We chose VDR as the representative technical probe based on a survey of 13 read-aloud and screen-reader applications regarding how well they support features for auditory skimming (e.g., navigation, speech rate). We provide the full comparison table, in which VDR stands out, as supplementary material.

The participants were allowed to use the temporal navigation features of VDR including play/pause, scrolling, jump forward or backward by 15 seconds, double tapping to start reading from a place, and page jumps. VDR also supports semantic navigation that takes sentence, paragraph, and section boundaries as delimiters but we did not incorporate them in our study as they did not work reliably for different PDFs that did not delimit texts homogeneously.

Procedure

Participants first answered demographic questions (age, gender, discipline). Then we demonstrated how to use the features of VDR and asked them to try them out. To simulate the commuting context, we projected a big screen in front of the participant and played a video from the perspective of a passenger (in Figure 2). To simulate ambient noise we kept the audio volume of the video high. Time allocations for each reading task were determined by the results from pilot tests. We gave participants 8.5 minutes for A1, 8 minutes for A2, and 4.5 minutes for each of A3 and A4. After each reading, they answered the questions on their perceived level of comprehension. Participants performed tasks in either original layout or plain text layout first. After finishing two tasks in each layout condition, they answered a NASA-TLX survey for the tasks in the condition.

Data collection and data analysis

We quantitatively measured: (1) NASA-TLX six scale weighted measurement for workload, and (2) perceived level of comprehension—a 5-point Likert scale measuring how well they feel about their performance for each skimming objective. For analysis, we conducted a paired t-test to analyze the workload data. We used Aligned Rank Transform [61] and then ANOVA to analyze their perceived level of comprehension.

We also collected qualitative data such as observation notes, video recording of the tasks, and audio recording of the post-task interviews. The semi-structured post-task interview was composed of: (1) factual questions asking for the gist and structural organization of the task material, (2) conventional

UI/UX questions pertaining to what worked well and what didn't, and (3) questions aiming to deepen our understanding of what we observed. For analysing the qualitative data, we conducted an in-vivo coding [15] followed by memoing and later more abstract coding. Codes and memos were shared within the research team, and refined through joint discussions.

Findings

The central theme of the qualitative results, which stood out more strongly than any other result in our study, is that ***the pull to look at the device is very strong.*** Participants almost seemed compelled at times to look down when skimming. Analysis of the video footage shows that participants were visually "scanning", which they confirmed in the post-task interview. They indicated that they were visually foraging parts of text where they believed the gist of the paper's content might be (e.g., end of Introduction, and places marked by phrases such as "In this paper" or "Our contributions are").

This tendency to look at the document was so strong that four participants would not follow the study protocol that requested them to minimize their visual attention to the screen. They read along most of the time, meaning that they visually read the text while just following the auditory narration as a secondary information channel. Individual preferences seemed to play into such behavior as some expressed themselves as a "visual person" and explained that "[listening to] audio is not learning." (P10) We discarded their quantitative data to keep homogeneity in our analysis (below) but kept their qualitative data, as their responses add accounts for why the pull to look is so strong.

Overall, the participants struggled considerably between using two different modes of reading: eyes-free vs. visual reading, which we had not anticipated. Skimming, as a particular type of reading behavior, requires frequent non-linear navigation and occasional glancing over to visual content, such as figure. However, the traditional read-aloud apps are primarily designed with the assumption of a linear consumption of audio. The extent of participants' difficulties with eyes-free motivates a strong need for a new auditory skimming app that supports a new mode of text consumption: *eyes-reduced* reading that flexibly blends auditory and visual modes of reading. We structure our remaining findings as concrete user needs and challenges for supporting skim-reading in an eyes-reduced way.

Non-linear navigation is hard. Participants were in need of non-linear navigation features that leverage the semantic structure of a document (e.g., sentences, paragraphs) so they won't feel lost right after jump, which is a frequent navigational pattern in skim-reading. They did not find *temporal* jumps (by 10s, 15s, and so on) useful, because *"It wasn't clear where 15 seconds leap would take me to. Move by sentences is more meaningful."* (P8) In general, participants reported being familiar with the structure of research papers and what to expect from them. *"I wanted to navigate. I wanted to see the results, see the summary, [go] to the suggestions. But it was not easy to navigate to them. You have to scroll and find."* (P4) Participants wanted to listen to the first one or two sentences of a paragraph and skip to next paragraph. However, they often did not skip because *"double tapping the next paragraph was not working always."* (P8) In addition, it required visual input from them to make a selection. Navigation was difficult with plain text layout because *"there was no structure and everything looked similar"* (P16) and participants had to *"scroll and scroll until at some point I became frustrated."* (P9)

Listening and navigating at the same time is difficult. The traditional app keeps narrating the text even after the user starts the navigational interaction (e.g., scroll to find where to jump into). Participants found this simultaneous stimulus cognitively taxing. It is known that listening and short-term memory is competing for the shared cognitive resources [49]. Their work-around was to pause the audio before navigating, although cumbersome.

Some types of content are difficult to understand by listening. Some participants preferred *looking down* at a figure or a table referenced in the text, because they were visual by nature: *"It's useless to say Figure 2 when you can't actually see Figure 2."* (P7) Some of the *textual* content was also difficult to understand, even when the narration is reading it verbatim. Such items include complex numbers, and acronyms: *"[It was] difficult to understand when the voice is telling p value is 0.3."* (P7) *"[It was] wordy, the stat reports are about numbers, and hearing the voice reading the numbers was not useful."* (P13)

Original layout better supports structural understanding. Participants rated that they understood the structure of the article better with the original layout ($M = 3.81, SD = 0.44$) than the plain text layout ($M = 2.38, SD = 0.35$); $F(1,15) = 28.31$, $p < .001$. They reported that *"plain text makes it difficult to see the actual structure of the article, ... for skimming, original layout was much better."* (P17) While doing the task, workload was significantly higher in plain text layout ($M = 76.67, SD = 11.87$) than in original layout ($M = 63.97, SD = 15.22$); $t(15) = -4.42$, $p < .001$.

Dynamic speech rate control is needed but inaccessible. Participants reported that they preferred slow narration for important content and faster narration for trivial content. Such a selective reading preference is similar to what is reported in the literature on visual skim-reading behaviors [19]. However, participants often could not increase or decrease the speech rate, even when they wanted to, because the rate control was buried deep in the menu, not accessible in an eyes-free manner. *"But since every time I have to touch it and decrease and increase the word speed, this deviated me from what I was hearing. That's why I skipped increasing the speed later."* (P12) Moreover, setting the base-rate that suits individual preference was another barrier, because the unit of speech rate "words per minute" was not intuitive for them to estimate how fast (or slow) it would be and they had to listen to the audio to figure out which speed to set.

There are individual differences in preferred reading order. We found varying reading orders for research papers. This order seemed predetermined by a person, something they learned and refined over their experience of reading articles of a similar genre. For example, the order of Abstract-Introduction-

Conclusion was very popular. Some would go to Discussion when time permits. Only P16 and P19 said they would be interested to see the figures first. For Pew reports, which were relatively new to many, they either: (1) listened chronologically, (2) visually read section, sub-section titles, and listened to one or two paragraphs, or (3) listened to the first paragraph and last paragraph of each section.

Document formatting and article type impact comprehension. When it comes to understanding the topic, the participants perceived to have better understanding: (1) while using original layout ($M = 4.16, SD = 0.13$) than plain text layout ($M = 3.53, SD = 0.31$); $F(1,15) = 7.147$, $p < .05$, and (2) when reading research papers ($M = 4, SD = 0.35$) than Pew reports ($M = 3.69, SD = 0.53$); $F(1,15) = 5.43$, $p < .05$. As for the key takeaway, likewise, they perceived to have better understanding (1) while using original layout ($M = 3.31, SD = 0.80$) than plain text layout ($M = 2.63, SD = 0.53$); $F(1,15) = 11.84$, $p < .05$, and (2) when reading research papers ($M = 3.44, SD = 0.62$) than Pew reports ($M = 2.5, SD = 0.35$); $F(1,15) = 14.27$, $p < .05$.

People desire to create annotations for later consumption. Part of the motivation for skimming is to decide whether the document needs detailed reading later. To this point, our participants mentioned their interest to create annotations such as, highlighting, putting stars, and bookmarking. Another motivation is to mark which content is important and go back to it sometime later.

To summarize our findings from this study, auditory skimming required 'eyes-reduced' reading rather than entirely eyes-free. Participants blended visual (for content that is predominantly visual, such as images) and non-visual modes of reading in order to skim using audio. Specifically, non-linear navigation was key to successful skimming, but it is not well supported.

DESIGN GUIDELINES FOR EYES-REDUCED SKIMMING

We triangulated implications from multiple data sources to develop a set of robust and comprehensive design guidelines. The data sources include: (1) the results of our Needs-Finding Study, (2) existing theoretical and empirical studies of skim-reading and eyes-free interactions, and (3) existing guidelines for creating voice narrations and designing VUI.

DG-1. Provide ways to navigate the structure of the article in a non-linear fashion and to localize the current position. Non-linear and structural navigation features should be readily available to the user on demand, similar to using a Table of Contents. The primary support required here is the ability to jump to sections and to learn the current position in the document which is also supported by the findings in [9, 35].

DG-2. Provide semantic and spatial navigation instead of temporal navigation. As suggested by the findings in [4, 19, 35], supporting semantic navigation enables users to leverage meaningful words and structural markers (e.g., chapters, paragraphs) as navigational cues. Moreover, informed by our Needs-Finding Study and Hyland and Tse's discourse analysis on academic articles [26], we recommend supporting discourse marker jumps to facilitate skimming since they also act as important cues when skimming visually. Hyland and Tse illustrated two major types of discourse markers: interactive and interactional. Interactive markers are those that the writer uses to manage information flow as a guide to the reader, such as 'finally', 'to conclude', 'in this paper', and 'see Figure 1'. Interactional markers are those that convey the author's perspective towards the propositional information. For example, 'it is clear that', 'note that', and 'our contributions'. In addition, supporting spatial navigation (e.g., vertical scroll) helps the user to leverage visual markers in documents as navigational cues at times when the user momentarily glances at the screen. On the other hand, we found that temporal navigation is not desirable; our participants were unsure where the jump would take them to.

DG-3. Pause narration when the user is navigating. Narration should be paused when the user is navigating, such as scrolling or dragging. Our Needs-Finding Study showed that users find it difficult to follow the narration while thinking about where to go next at the same time.

DG-4. Provide ways to adjust speech rate dynamically. The user should be able to adjust the speech rate dynamically [6, 35]. Further, our Needs-Finding Study suggested that the user should be able to control the speech rate without stopping the narration currently underway or having to go to sub-menu level hierarchy. Moreover, the speech rate should be presented in a relative scale (e.g., 1.0 for normal speed, 2.0 for 2× the normal speed) instead of WPM which is not intuitive.

DG-5. Provide ways to refer back to text content from the narration and vice versa. Showing a visual indicator of a spatial point of reference (i.e., the text that the app is currently narrating) can help the user to refer back to the text content from the narration. There are two ways to achieve this. First, the indicator remains static on the screen, and the content 'flows' from behind. The user should be able to decide which text to narrate by directly manipulating the flow of content. Second, the indicator can be moved across to the content. The user should be able to move the indicator and start narration from anywhere in the document. Existing apps provide mixed support for the two mechanisms; e.g., while VoiceOver only supports the latter, VDR users can toggle between the two.

DG-6. Diverge from verbatim narration for specific types of text to enhance listening comprehension. For a better listening experience and comprehension, the speech synthesizer may narrate texts in a non-verbatim fashion. For example, break a long list down to groups of 3 or 4 items, narrate statistical values following verbal reporting conventions, expand abbreviations ('e.g.', 'i.e.'), insert breaks in long sentences to help the user absorb information, and avoid narrating footnotes and citations in the text [17, 21]. Moreover, the system should add *context* to help the user to focus. For example, announcing which section is going to be narrated next.

DG-7. Provide auditory or haptic feedback as non-visual navigation cues. Provide auditory or haptic feedback to guide non-visual navigation [14, 35]. For example, auditory feedback can be given between the end of a paragraph and the beginning of a new paragraph, to nudge the user to pay attention to structural boundaries in the content.

DG-8. Support opt-in visual engagement. Findings from our Needs-Finding Study indicate that there should be support for opt-in processes when switching context from the main text to figure/table captions. Additionally, when the system narrates a complex number that is difficult to understand by listening, the system should provide enlarged numeric text on the screen for easy visual access.

DG-9. Support unimanual interactions. Unimanual support is required in other situational impairment contexts, such as encumbrance [45]. Menus and buttons should be reachable with the hand holding the device, assuming that the user will be primarily interacting with one hand.

DG-10. Support individual differences in skimming strategies. The user should be allowed to pre-set a custom reading order in which the app will visit document elements automatically. This can happen when the reader is familiar with the genre of the document. We call this *passive skimming*. Furthermore, there should be support for choosing which part to listen to on the fly. This can happen when the user is not familiar with the document. We call this *active skimming*.

DG-11. Support annotation creation and consumption. The user should be able to create within-text telegraphic annotation [36] such as highlights, bookmarks, and stars. The user should be able to navigate between annotations. An advanced level of support would be to create speech annotations which are known to be useful in an eyes-free context [31]. At the same time, there should be support to consume these annotations effectively [22].

DESIGN OF SKIMMER

We embodied a subset of the guidelines for designing eyes-reduced skimming (DG-1 to 9 and partially DG-10) into a novel read-aloud mobile app. Skimmer features navigational support for auditory and visual modes of reading, non-linear navigation, and auditory and haptic cues. DG-11 was not considered in our design because the annotation tasks are only relevant to skimming when the user intends to revisit the document for a full read later on, and hence, is beyond the scope of the current stage of our work.

We designed and developed Skimmer through an iterative design process. Our first step was to create a conceptual design of a basic spatial layout and gesture interactions. In terms of layout, both the original layout and the plain text layout had different merits in skimming. Thus, as an optimal solution, we built Skimmer to support HTML-based e-publication format as a source document type (e.g., an HTML version of a paper in the latest CHI Proceedings), which marries the merits of both layouts. As the e-pub scene is moving from PDFs to such flexible document types for enhancing accessibility and device-compatibility, we consider using the HTML version of a document to be a future-facing decision. Note that an HTML-based document can be browsed in a single page by scrolling, rather than paginated.

The touch gesture vocabulary in Skimmer was fine-tuned through rounds of iterative design to support optimal unimanual interactions (DG-9). To get the right design [54] the lead author iterated through 3 versions of the gesture sets, tested with a total of 8 pilot users during in-depth brainstorming and design-review sessions over 3 months.

Here, we elaborate how Skimmer supports eyes-reduced skimming for SIs and our rationale behind its design.

Basic navigation for eyes-reduced skimming

Skimmer predominantly supports eyes-free skimming. The user can skip sentences or re-listen to a missed passage with a lightweight, eyes-free tap gesture on the right or left side of the screen respectively (Figure 1(a)). Paragraph navigation is done by up or down vertical flicking gestures. This is because scrolling is so innate in people and HTML-based documents flow vertically, so Skimmer aims to promote positive transfer from visual scrolling to eyes-free gestural interactions. When the spatial point of reference progresses to the next sentence or jumps to a remote one, the content automatically scrolls vertically so that the reference position is always located right at the middle of the screen. This makes localization (DG-5) as easy as glancing at the screen to check which text the indicator is pointing at (the red arrow in Figure 1). Abiding by the notion of eyes-*reduced*, Skimmer also supports visual navigation by traditional vertical scrolling (DG-2). Skimmer pauses the audio as soon as the user puts their touch on the screen for navigational action (DG-3).

Easy-to-understand audio narration

We designed Skimmer's audio narration in a way that helps listeners better understand complex and difficult texts (DG-6). First, written text and structural information are narrated in two different voices to help distinguish meta-information from verbatim content. This is done by using a low-pitch voice for narrating written text and a higher-pitch voice for narrating meta-information. The meta-information is for providing context to the reader, such as "Section 1. Introduction", where "Section" is the meta-information. Second, Skimmer breaks down long lists of items or bullets in groups of 3 or 4 items at a time, to help memory retention. Finally, Skimmer reads numbers in a rounded format (such as $p < 2.2 \times 10^{-16}$ narrated as $p < .001$) because complex numbers are hard to understand just by listening. Additionally, Skimmer provides an enlarged text overlay of the number (Figure 1(b)). Another example is the emphasis on discourse markers, calling the user's attention to important information: when the original text is ("In this paper, we study Snapchat ..."), it is rephrased to *"<emphasis level="strong">In this paper,</emphasis> we study Snapchat ..."*.

Discourse marker navigation

To help the user find and navigate to important nuggets of information, Skimmer supports selective jumps between various discourse markers (DG-2). When the user taps the discourse marker navigation button on the bottom right (or left, see Figure 1(a)), Skimmer first moves the play-head to the next (or previous) adjacent discourse marker, reads the discourse marker (e.g., "In this paper"), and then narrates the full sentence from the beginning of the sentence that contains the discourse marker. Both discourse markers and full sentences are narrated by the same low-pitch voice. As feedback to help the user disambiguate discourse marker narrations from the normal narration, Skimmer plays a subtle ambient earcon (e.g., cricket sound) as background audio.

Overview to support understanding document structure

To support structural glancing and navigation over high-level structure of the document (DG-1), Skimmer features a table-of-content style *Overview* page apart from the main document view (Figure 1(c)). All content in the Overview is accessible eyes-free, as the allowed gestures between the main document view and the Overview are consistent. This also supports individual differences in skimming strategies, in part, by providing opportunities to seamlessly choose which part of the document to listen to (DG-10).

Haptic cues to support visual opt-in

Skimmer's design for eyes-reduced reading nudges the user when the spatial point of reference encounters visual content. When narrated content includes a figure or table referred to in the text, Skimmer gently informs the user with haptic feedback so that they can look at the visual content (DG-7, DG-8). Skimmer facilitates navigating back and forth between the main text and the visual content on-demand; note that we designed the transition for visual peeking to be *opt-in* to allow the user to choose to continue to listen to the text or to jump to looking at the visual content (see Figure 1(d)). The haptic pattern was carefully chosen; from Seifi et al.'s [48] library of vibration patterns, we tested "attention-catching" and "get ready" to explore vibration patterns. From among five vibration patterns, we selected the one that was found to be the most pleasant by our pilot participants.

Auditory cues to signify critical skimming moments

Inspired by SeeReader [14], Skimmer leverages a limited set of distinctive earcons to provide subtle auditory feedback (DG-7). We consciously limited the number of total distinctive earcons to be under 6, as we found that incorporating a number of earcons can cause overlap of aural characteristics between them, which in turn confuses the user. Hence, we used earcons, a scarce design resource due to the limited number that can be effectively leveraged, to signify the status changes at critical moments of skimming, such as switching between sentences, paragraphs, and discourse markers, and navigating to the Overview page—thus giving the user clear feedback that Skimmer has processed their navigational gestures.

Accommodating individual differences in listening rates

Skimmer accommodates individual differences by enabling the user to dynamically adjust the speech rate through an eyes-free tap gesture (DG-4). Skimmer provides two buttons at the top, similar to the discourse marker buttons in style and size (Figure 1(a)). Although the top of the screen is often hard to reach, we made a design trade-off by prioritizing discourse marker navigation, an oft-used navigational feature, over the speech rate changes.

In summary, our iterative design process yielded a novel interface that embodies the eyes-reduced skimming guidelines into a medium-fidelity prototype system.

Implementation

Skimmer was implemented as a Javascript-based webapp running on a 6.4-inch mobile phone (Samsung A50). We generated audio narrations using a Google Wavenet-based TTS engine, which was evaluated by our pilots to be easier to understand than other options (e.g, Amazon Polly and IBM Bluemix) with respect to the metrics proposed in [58].

PRELIMINARY EVALUATION OF SKIMMER

We conducted a preliminary evaluation of Skimmer. It was a structured observation on a public bus where participants experienced both Skimmer and VDR (also used in our Needs-Finding Study). We used VDR as a reference point to ground participants' audio skimming experience, not with an expectation for it to be evaluated "head to head" (as VDR was not designed for auditory skimming). Our primary evaluation goal was to assess *qualitatively* how the participants experienced the unique features of Skimmer in a realistic SI setting. With this decision, we traded-off against running a controlled comparative study, which would have required a lab study that doesn't offer the same ecological insights.

Method

Six graduate students (3F/3M) from various disciplines took part (none from our Needs-Finding Study). For the task, we asked participants to skim two documents in an "eyes-reduced" manner, one document with each application. The motivation, context, and protocol were similar to Needs-Finding Study. As reading materials, we chose two articles from the CHI 2019 Proceedings, both related to online sharing and social media (hereafter referred to as A1 [23] and A2 [8]). The articles are isomorphic in style, structure, and difficulty (on the Flesch-Kinkaid Reading Ease scale [30]). Each participant experienced both articles; orders of both articles and apps was counterbalanced. Times allotted for skimming A1 and A2 was 9 and 8 minutes respectively, due to differences in length (10,275 vs. 7,745 words). The procedure also echoed that of Needs-Finding Study, except that this study was conducted on the bus. For each app, we first demonstrated it to the participant and let them play with it to become comfortable using its features. The lead investigator accompanied each participant on the bus, observed them completing their tasks, and took notes. We used a two-way audio splitter so that both the participant and the investigator could listen to the audio output from the app individually. After each app-article condition, the participant got off the bus, answered a questionnaire on their article comprehension, and then sat for an interview (audio-recorded).

Findings

All 6 participants completed the task two times, i.e., they skim-read the two articles, one with each app. Quantitatively, they performed comparably on the reading comprehension questionnaires for both apps (data is given in the supplemental material). We focus on the qualitative comparison here. We note that, in the interview, three participants reported, without prompting, that they typically experience motion sickness in a moving bus.

Skimmer can be used eyes-reduced, whereas VDR requires near constant visual attention. It was very clear in our observation that participants were able to largely not look at the screen using Skimmer. There were intermittent instances where participants glanced at the screen (1) in the Overview page to jump between sections, (2) in the Article page to peek at intriguing keywords, to react to (unexpected) earcons (e.g., appearance of a new section after the next paragraph gesture), and to respond to a haptic nudge when a figure or table was referred in the text. P2, P4, and P5 looked at the screen in the Results section, skipping by paragraphs, often stopping for a brief moment "to find interesting results quickly". The comparison to VDR was black and white: participants spent the majority of their time visually scrolling *"all the way down to find Conclusion. I don't have any other way of doing that other than looking at the screen. But it was really easy with this application (Skimmer)."* (P2)

Users appreciate Overview the most. The Overview was called out as the most useful feature in Skimmer: "key difference between the two apps." With Skimmer, participants frequently opened the Overview page, scrolled through the list of sections and sub-sections, glanced over the screen intermittently to find a section, and then jumped to that section largely eyes-free. At one moment, P3 and P4 were tapping instead of swiping to the left because they "forgot how to go back [to the Article/text]", which also prompted them to look at the screen. In contrast, section-level jumps in VDR prompted heavy visual attention and frequent visual scrolling.

Auditory feedback and haptic feedback help to re-focus. While the intended purpose of the auditory and haptic cues was to subtly signal an interaction (e.g., jump to next sentence) or signal an affordance (e.g., a figure that can be seen), they also helped participants "to situate themselves and come back" when they lost focus. However, some participants found it difficult to notice some of the subtle earcons.

Individual differences in navigation preferences are supported. P2 jumped from the Conclusion to the last sentence of the Abstract by using a combination of section, paragraph, and sentence jumps as "shortcuts". P1, P3, and P6 used sentence jumps frequently because "it was faster", while P4 and P5 frequently used dragging for sentence jumps and occasionally flicked for paragraph jumps because "scrolling felt natural".

Skimming a complex number remains challenging for those who tried it. With the exception of P5 and P6, other participants did not consider numbers important while skimming. P5 looked down upon numbers in the Results section and visually read with both VDR and Skimmer because *"it was difficult to follow numbers, and defeated the purpose of audio skimming."* Both P2 and P5 resonated that *"numbers should come with context, for example, 1500 participants"* instead of just "1500".

Users appreciate the quality of narration and multiple voices. Participants were impressed by the quality of narration in Skimmer, mentioning that it was *"soft and comfortable to listen to, unlike the stiff voice in the other (VDR)".* In contrast, P5 mentioned that she liked VDR's narration because it was *"natural, research papers are dry, you don't need emotion here, a robotic voice is also okay."* Skimmer's use of multiple voices was "unique" but little noticed.

Discourse markers are useful, but users need more exposure to the feature. P3 and P5 used the discourse marker jump and described it as a "great tool" that is "helpful" to find important information. The others didn't use the feature despite all of them experiencing it during training. The non-users commented that they "needed more trust." P1 said, *"I may need to experience it just by myself alone before I know what really is in there".* In contrast, with VDR, participants "zoom-in, [and] spot discourse markers" which is an interaction pattern similar to what we observed from our Needs-Finding Study.

Figures/Tables are mostly ignored, but participants appreciated the idea of a haptic nudge. P1 and P3 used the haptic nudge feature and described it as *"a great way of reminding me that something is happening here."* (P3) P3 added that *"I don't do audio reading because I am not good at keeping attention. But such vibrations are really useful."* The rest of the participants ignored the opt-in feature, citing "figures/tables [are] not important in skimming". By contrast, while using VDR, P1 and P3 momentarily stopped at a figure/table while scrolling visually, zoomed-in and out, and then moved on.

Participants appreciated Skimmer's design concept. The participants made many encouraging comments about Skimmer's eyes-reduced design approach for addressing the challenge of on-the-go skimming. P5 said that "[there is] lots of reading for a PhD student, not a terrible idea to get gist for some [...] super helpful in skimming qualitative papers." P2 suggested that it could be used for visual and auditory readings in parallel: "read-along for listening and following words". P6's comment highlights the key research motivation of this study; when using VDR, she felt nauseated and had to stop momentarily. By contrast, she did not report feeling nauseated when using Skimmer and was able to complete the task eyes-reduced.

DISCUSSION

We reflect on our findings and their implications for design.

We have made concrete progress towards eyes-reduced skimming. Our goal for this research was to design an interface that would allow people to skim documents in an eyes-reduced way. Given the complexity of skimming as a form of reading [18, 19] that frequently mixes non-linear reading and glancing, supporting it via the audio channel is a non-trivial design objective. Especially for an on-the-go setting, where factors of distraction and motion sickness are prevalent, it poses more challenges. In our Preliminary Evaluation *on the bus*, we could only ask participants to try not to use their eyes, but we could not force them to not look. We were pleasantly surprised at the extent to which participants did in fact not use their eyes while using Skimmer, which was not possible with the traditional read-aloud application. Further, our participants were positive about the overall design concept of Skimmer, with 5 out of 6 of them indicating that they would continue using the app for skimming documents. This suggests that we have indeed made positive progress towards our design goal, that the Skimmer design is on the right track.

Skimmer is a design snapshot of an artefact in its evolution. Even though our Preliminary Evaluation was promising, there are opportunities to revise the design guidelines, improve usability, and address the gaps in our design for supporting eyes-reduced interaction. For example, we can revise DG-8 to recommend inclusion of context rather than highlighting only the number. We only explored some design dimensions in our work, such as numbers and lists. Further, new elements such as a means to navigate to and from metadata (headers, footers, and references) will need to be integrated such that they do not make the interaction overly complex.

Eyes-reduced is a new concept that bridges full-visual and eyes-free interaction. Our work has introduced the concept of eyes-reduced design—it is intended to capture designs that can be used primarily eyes-free but, importantly, do not preclude the user from using their eyes, when they want. This can be differentiated from design that is created with the intention of being used with one's eyes but also happens to support some degree of eyes-free interaction. For example, a person using a mobile YouTube app can select a video while looking and then, to some extent, look away while walking, toggling the pause/play to listen to the video. This is challenging because the pause/play tap region is small, specified by a visual icon. Thus, while it can be used eyes-free for some interactions, it was not seemingly designed with that intention.

Our eyes-reduced design for skimming can serve as an intermediary between designs for full-visual skimming and those for eyes-free skimming by visually impaired (VI) people. We believe that our study is an important step towards pushing for eyes-free by means of eyes-reduced. In general, designing for people with disabilities is known to also benefit those experiencing situational impairment (SI) [60]. In that sense, some of our design guidelines were informed by guidelines established for VI, as outlined in our Guidelines section. Determining which guidelines from the VI context are transferable to SI is a non-trivial contribution of our work.

Fully eyes-free skimming is the next main target of our research. Our goal from the outset of this research was that Skimmer support both the needs of people experiencing SI as well as those who have visual impairments. We made the choice to start our design process by targeting SI, as we reckoned it may be the more difficult design problem—people in SI still have an option to use their eyes (hence eyes-reduced) when they deem it necessary. This goes back to the point made above on how can we realize eyes-free design by way of eyes-reduced. Some design insights that take us closer to eyes-free design are: text narration needs to go beyond simply narrating the given text with verbatim speech. There is room for improvement not only in what is said (i.e., content) but also how it is said (e.g., paralingual cues, including voice modulation, varying pitch, tone, and varying speech rate) for better comprehension.

Some other extensions to support fully eyes-free are immediately obvious, such as reading out figure descriptions (i.e., alt text). After incorporating those, our work will then continue with systematically analyzing needs of VI individuals and design for them. Evaluation will undoubtedly reveal that some of Skimmer's design elements need adjustment to support that population. Upon iterating on the design, we will loop back to evaluating with people experiencing SIs so as to ensure that our design is maximally supporting both eyes-reduced and eyes-free.

Different contexts pose different forms of SIs necessitating varying types of support. We only explored a context where people could not fully use their *visual* attention because of potential motion sickness. However, different environments/contexts result in SIs that vary in degree. For example, one riding a commuting bus can occasionally glance at their mobile, but it will be nearly impossible for one driving a vehicle to use their eyes, even momentarily.

Also, a specific impairment context may pose challenges in using interaction modalities other than vision. Zeleznik et al., documented the 'sandwich' problem when a user is holding an object in one hand, which impairs one's manual interaction capability [65]. We wonder about other types of SIs that variably affect different modalities. For example, what about when people can see, but cannot touch the device (e.g., browsing recipes when your hands are oily or wet)? Interfaces that address specific kinds of SI is a promising direction.

LIMITATIONS

There are limitations to our work. We focused on structured, academic/professional documents. It remains to be seen how well Skimmer will work with documents that have a lot of visual content or ones with many links that require cross-document navigation, thus beyond non-linear. Further, our evaluation of Skimmer only involved 6 graduate students, with a single exposure. A larger sample from a more varied population, together with a longer exposure time to Skimmer, will be needed to more deeply assess the long-term, ecological efficacy of the design elements.

CONCLUSION

In this work we tackled the design challenge of supporting skimming of structured documents on the go. Based on our Needs-Finding Study, we conceptualized eyes-reduced design as a viable solution approach to the problem, generated design guidelines for eyes-reduced skimming, iterated extensively on the design based on the guidelines, and then implemented Skimmer, a proof-of-concept prototype. Through a preliminary evaluation, we learned that people could use Skimmer eyes-reduced while riding the bus, only looking down at the app infrequently, as they navigated a document largely eyes-free, listening to the content.

We are conducting our work within the framework of inclusive design [16], with planned next steps to involve visually impaired people so that we can push more strongly on eyes-free interaction. This will lead us one step closer to achieving a system that people who are experiencing impairments (either situationally or more permanently) can use to skim documents.

ACKNOWLEDGEMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Mortimer J Adler and Charles Van Doren. 2014. *How to read a book: The classic guide to intelligent reading*. Simon and Schuster.

[2] Faisal Ahmed, Yevgen Borodin, Yury Puzis, and I. V. Ramakrishnan. 2012a. Why Read if You Can Skim: Towards Enabling Faster Screen Reading. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '12)*. ACM, New York, NY, USA, Article 39, 10 pages. DOI: http://dx.doi.org/10.1145/2207016.2207052

[3] Faisal Ahmed, Yevgen Borodin, Andrii Soviak, Muhammad Islam, I.V. Ramakrishnan, and Terri Hedgpeth. 2012b. Accessible Skimming: Faster Screen Reading of Web Pages. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 367–378. DOI: http://dx.doi.org/10.1145/2380116.2380164

[4] Faisal Ahmed, Andrii Soviak, Yevgen Borodin, and I.V. Ramakrishnan. 2013. Non-visual Skimming on Touch-screen Devices. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI '13)*. ACM, New York, NY, USA, 435–444. DOI: http://dx.doi.org/10.1145/2449396.2449452

[5] Monica Anderson and Jingjing Jiang. 2018. Teens, Social Media & Technology 2018. Pew Research Center. (13 May 2018). Retrieved September 18, 2019 from https://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/.

[6] Barry Arons. 1993. SpeechSkimmer: Interactively Skimming Recorded Speech. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology (UIST '93)*. ACM, New York, NY, USA, 187–196. DOI: http://dx.doi.org/10.1145/168642.168661

[7] Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: Subtle and Eyes-free Mobile Input with a Magnetically-tracked Finger Ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2043–2046. DOI: http://dx.doi.org/10.1145/1978942.1979238

[8] Frank Bentley, Katie Quehl, Jordan Wirfs-Brock, and Melissa Bica. 2019. Understanding Online News Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 590, 11 pages. DOI: http://dx.doi.org/10.1145/3290605.3300820

[9] Sara Bly and Catherine C Marshall. 2005. Turning the page on navigation. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*. IEEE, 225–234.

[10] Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley. 2011. Experimental Analysis of Touch-screen Gesture Designs in Mobile Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 403–412. DOI: http://dx.doi.org/10.1145/1978942.1979000

[11] Danielle Bragg, Cynthia Bennett, Katharina Reinecke, and Richard Ladner. 2018. A Large Inclusive Study of Human Listening Rates. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 444, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3174018

[12] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal Eyes-free Interaction Techniques for Wearable Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 473–480. DOI: http://dx.doi.org/10.1145/642611.642694

[13] Moritz Büchi, Natascha Just, and Michael Latzer. 2017. Caring is Not Enough: The Importance of Internet Skills for Online Privacy Protection. *Information, Communication & Society* 20, 8 (2017), 1261–1278.

[14] Scott Carter and Laurent Denoue. 2009. SeeReader: An (Almost) Eyes-Free Mobile Rich Document Viewer. *CoRR* abs/0909.2185 (2009). http://arxiv.org/abs/0909.2185

[15] Kathy Charmaz and Linda Liska Belgrave. 2007. Grounded theory. *The Blackwell encyclopedia of sociology* (2007).

[16] P John Clarkson, Roger Coleman, Simeon Keates, and Cherie Lebbon. 2013. *Inclusive design: Design for the whole population*. Springer Science & Business Media.

[17] Michael H Cohen, Michael Harris Cohen, James P Giangola, and Jennifer Balogh. 2004. *Voice user interface design*. Addison-Wesley Professional.

[18] Geoffrey B Duggan and Stephen J Payne. 2009. Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied* 15, 3 (2009), 228. http://dx.doi.org/10.1037/a0016995

[19] Geoffrey B. Duggan and Stephen J. Payne. 2011. Skim Reading by Satisficing: Evidence from Eye Tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1141–1150. DOI: http://dx.doi.org/10.1145/1978942.1979114

[20] Juliane Franze, Kim Marriott, and Michael Wybrow. 2014. What Academics Want when Reading Digitally. In *Proceedings of the 2014 ACM Symposium on Document Engineering (DocEng '14)*. ACM, New York, NY, USA, 199–202. DOI: http://dx.doi.org/10.1145/2644866.2644894

[21] David Goldberg. 2019. The Voice Actor's Performance Guidebook. (2019). Retrieved September 19, 2019 from https://www.edgestudio.com/voice-over-performance-guidebook.

[22] Jonathan Grudin. 1988. Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces. In *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work (CSCW '88)*. ACM, New York, NY, USA, 85–93. DOI:http://dx.doi.org/10.1145/62266.62273

[23] Hana Habib, Neil Shah, and Rajan Vaish. 2019. Impact of Contextual Factors on Snapchat Public Sharing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 26, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300256

[24] Terje Hillesund. 2010. Digital reading spaces: How expert readers handle books, the Web and electronic paper. (2010).

[25] Juan David Hincapié-Ramos and Pourang Irani. 2013. CrashAlert: Enhancing Peripheral Alertness for Eyes-busy Mobile Interaction While Walking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3385–3388. DOI: http://dx.doi.org/10.1145/2470654.2466463

[26] Ken Hyland and Polly Tse. 2004. Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics* 25, 2 (06 2004), 156–177. DOI: http://dx.doi.org/10.1093/applin/25.2.156

[27] Christopher Ingraham. 2016. The astonishing human potential wasted on commutes. (25 February 2016). Retrieved September 18, 2019 from https://www.washingtonpost.com/news/wonk/wp/2016/02/25/how-much-of-your-life-youre-wasting-on-your-commute/?noredirect=on.

[28] Shaun K. Kane, Jacob O. Wobbrock, and Ian E. Smith. 2008. Getting off the Treadmill: Evaluating Walking User Interfaces for Mobile Devices in Public Spaces. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '08)*. ACM, New York, NY, USA, 109–118. DOI: http://dx.doi.org/10.1145/1409240.1409253

[29] Dohyung Kim, Gahgene Gweon, and Geehyuk Lee. 2016. SATS: Structure-Aware Touch-Based Scrolling. *ETRI Journal* 38, 6 (2016), 1104–1113. DOI: http://dx.doi.org/10.4218/etrij.16.2716.0017

[30] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. (1975). https://stars.library.ucf.edu/istlibrary/56

[31] Antti Konttila, Marja Harjumaa, Salla Muuraiskangas, Mikko Jokela, and Minna Isomursu. 2012. Touch n'Tag: Digital annotation of physical objects with voice tagging. *Journal of Assistive Technologies* 6, 1 (2012), 24–37. https://doi.org/10.1108/17549451211214337

[32] Esther Landhuis. 2016. Scientific literature: Information overload. (2016). Retrieved January 02, 2020 from https://www.nature.com/naturejobs/science/articles/10.1038/nj7612-457a.

[33] Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5203–5214. DOI:http://dx.doi.org/10.1145/2858036.2858299

[34] Kevin A. Li, Patrick Baudisch, and Ken Hinckley. 2008. Blindsight: Eyes-free Access to Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1389–1398. DOI: http://dx.doi.org/10.1145/1357054.1357273

[35] Tonja Machulla, Mauro Avila, Pawel Wozniak, Dillon Montag, and Albrecht Schmidt. 2018. Skim-reading Strategies in Sighted and Visually-Impaired Individuals: A Comparative Study. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18)*. ACM, New York, NY, USA, 170–177. DOI: http://dx.doi.org/10.1145/3197768.3201535

[36] Catherine C. Marshall. 1997. Annotation: From Paper Books to the Digital Library. In *Proceedings of the Second ACM International Conference on Digital Libraries*. Association for Computing Machinery, New York, NY, USA, 131–140. DOI: http://dx.doi.org/10.1145/263690.263806

[37] Catherine C. Marshall and Christine Ruotolo. 2002. Reading-in-the-Small: A Study of Reading on Small Form Factor Devices. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries*. Association for Computing Machinery, New York, NY, USA, 56–64. DOI: http://dx.doi.org/10.1145/544220.544230

[38] Michael E Masson. 1982. Cognitive Processes in Skimming Stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8, 5 (1982), 400. https://doi.org/10.1037/0278-7393.8.5.400

[39] Bernard Merr. 2018. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. (21 May 2018). Retrieved September 18, 2019 from https://bit.ly/2krBBrv.

[40] Martez E. Mott and Jacob O. Wobbrock. 2019. Cluster Touch: Improving Touch Accuracy on Smartphones for People with Motor and Situational Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 27, 14 pages. DOI: http://dx.doi.org/10.1145/3290605.3300257

[41] Alexander Ng, Stephen A. Brewster, and John H. Williamson. 2014. Investigating the Effects of Encumbrance on One- and Two- Handed Interactions with Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1981–1990. DOI: http://dx.doi.org/10.1145/2556288.2557312

[42] Nosa Omoigui, Liwei He, Anoop Gupta, Jonathan Grudin, and Elizabeth Sanocki. 1999. Time-Compression: Systems Concerns, Usage, and Benefits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. Association for Computing Machinery, New York, NY, USA, 136–143. DOI: http://dx.doi.org/10.1145/302979.303017

[43] Jerome Pasquero, Scott J. Stobbe, and Noel Stonehouse. 2011. A Haptic Wristwatch for Eyes-free Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3257–3266. DOI: http://dx.doi.org/10.1145/1978942.1979425

[44] Martin Pielot, Anastasia Kazakova, Tobias Hesselmann, Wilko Heuten, and Susanne Boll. 2012. PocketMenu: Non-visual Menus for Touch Screen Devices. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 327–330. DOI: http://dx.doi.org/10.1145/2371574.2371624

[45] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of Situational Impairments During Interaction with Mobile Devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OZCHI '17)*. ACM, New York, NY, USA, 477–481. DOI: http://dx.doi.org/10.1145/3152771.3156161

[46] Bastian Schildbach and Enrico Rukzio. 2010. Investigating Selection and Reading Performance on a Mobile Phone While Walking. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 93–102. DOI:http://dx.doi.org/10.1145/1851600.1851619

[47] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *HCI international*, Vol. 2. 1298–1302.

[48] H. Seifi, K. Zhang, and K. E. MacLean. 2015. VibViz: Organizing, visualizing and navigating vibration libraries. In *2015 IEEE World Haptics Conference (WHC)*. 254–259. DOI: http://dx.doi.org/10.1109/WHC.2015.7177722

[49] Ben Shneiderman. 2000. The Limits of Speech Recognition. *Commun. ACM* 43, 9 (Sept. 2000), 63–65. DOI:http://dx.doi.org/10.1145/348941.348990

[50] Venkatesh Sivaraman, Dongwook Yoon, and Piotr Mitros. 2016. Simplified Audio Production in Asynchronous Voice-Based Discussions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1045–1054. DOI: http://dx.doi.org/10.1145/2858036.2858416

[51] Aaron Smith and Maeve Duggan. 2018. Crossing the Line: What Counts as Online Harassment? Pew Research Center. (04 January 2018). Retrieved September 18, 2019 from https://www.pewinternet.org/2018/01/04/crossing-the-line-what-counts-as-online-harassment/.

[52] Ainin Sulaiman, Noor Ismawati Jaafar, and Alireza Tamjidyamcholo. 2018. Influence of personality traits on Facebook engagement and their effects on socialization behavior and satisfaction with university life. *Information, Communication & Society* 21, 10 (2018), 1506–1521.

[53] Alexander Thayer, Charlotte P. Lee, Linda H. Hwang, Heidi Sales, Pausali Sen, and Ninad Dalal. 2011. The Imposition and Superimposition of Digital Reading Technology: The Academic Potential of e-Readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2917–2926. DOI: http://dx.doi.org/10.1145/1978942.1979375

[54] Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. 2006. Getting the Right Design and the Design Right. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1243–1252. DOI: http://dx.doi.org/10.1145/1124772.1124960

[55] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. 2006. Reading On-the-go: A Comparison of Audio and Hand-held Displays. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '06)*. ACM, New York, NY, USA, 219–226. DOI:http://dx.doi.org/10.1145/1152215.1152262

[56] Yolanda Vazquez-Alvarez and Stephen A. Brewster. 2011. Eyes-free Multitasking: The Effect of Cognitive Load on Mobile Spatial Audio Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2173–2176. DOI: http://dx.doi.org/10.1145/1978942.1979258

[57] Sunil Vemuri, Philip DeCamp, Walter Bender, and Chris Schmandt. 2004. Improving Speech Playback Using Time-compression and Speech Recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 295–302. DOI: http://dx.doi.org/10.1145/985692.985730

[58] Mahesh Viswanathan and Madhubalan Viswanathan. 2005. Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale. *Computer Speech & Language* 19, 1 (2005), 55 – 83. DOI: http://dx.doi.org/https://doi.org/10.1016/j.csl.2003.12.001

[59] Steve Whittaker, Julia Hirschberg, John Choi, Don Hindle, Fernando Pereira, and Amit Singhal. 1999. SCAN: Designing and Evaluating User Interfaces to Support Retrieval from Speech Archives. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*. Association for Computing Machinery, New York, NY, USA, 26–33. DOI:http://dx.doi.org/10.1145/312624.312639

[60] Jacob O Wobbrock. 2019. Situationally-induced impairments and disabilities. In *Web Accessibility*. Springer, 59–92.

[61] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

[62] Maryanne Wolf. 2018. Skim reading is the new normal. The effect on society is profound. (25 August 2018). Retrieved September 18, 2019 from https://www.theguardian.com/commentisfree/2018/aug/25/skim-reading-new-normal-maryanne-wolf.

[63] Bo Yi, Xiang Cao, Morten Fjeld, and Shengdong Zhao. 2012. Exploring User Motivations for Eyes-free Interaction on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2789–2792. DOI: http://dx.doi.org/10.1145/2207676.2208678

[64] Chen-Hsiang Yu and Robert C. Miller. 2011. Enhancing Mobile Browsing and Reading. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 1783–1788. DOI:http://dx.doi.org/10.1145/1979742.1979845

[65] Robert Zeleznik, Andrew Bragdon, Ferdi Adeputra, and Hsu-Sheng Ko. 2010. Hands-on Math: A Page-based Multi-touch and Pen Desktop for Technical Work and Problem Solving. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 17–26. DOI:http://dx.doi.org/10.1145/1866029.1866035

[66] Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. 2007. Earpod: Eyes-free Menu Selection Using Touch Input and Reactive Audio Feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1395–1404. DOI: http://dx.doi.org/10.1145/1240624.1240836