Exploring Interactive Sound Design for Auditory Websites

Lotus Zhang
University of Washington
Seattle, Washington, United States
hanziz@uw.edu

Jingyao Shao
University of Washington
Seattle, Washington, United States
jyaoshao@gmail.com

Augustina Ao Liu
University of Washington
Seattle, Washington, United States
liuao.uw@gmail.com

Lucy Jiang
University of Washington
Seattle, Washington, United States
lucjia@cs.washington.edu

Abigale Stangl
University of Washington
Seattle, Washington, United States
hanziz@uw.edu

Adam Fourney
Microsoft Research
Redmond, Washington, United States
adamfo@microsoft.com

Meredith Ringel Morris
Microsoft Research
Redmond, Washington, United States
merrie@microsoft.com

Leah Findlater
University of Washington
Seattle, Washington, United States
leahkf@uw.edu

ABSTRACT

Auditory interfaces increasingly support access to website content, through recent advances in voice interaction. Typically, however, these interfaces provide only limited audio styling, collapsing rich visual design into a static audio output style with a single synthesized voice. To explore the potential for more aesthetic and intuitive sound design for websites, we prompted 14 professional sound designers to create auditory website mockups and interviewed them about their designs and rationale. Our findings reveal their prioritized design considerations (aesthetics and emotion, user engagement, audio clarity, information dynamics, and interactivity), specific sound design ideas to support each consideration (e.g., replacing spoken labels with short, memorable audio expressions), and challenges with applying sound design practices to auditory websites. These findings provide promising direction for how to support designers in creating richer auditory website experiences.

CCS CONCEPTS

• **Human-centered computing** → **Auditory feedback**; *Web-based interaction.*

KEYWORDS

voice interaction, audio display, interaction design

ACM Reference Format:
Lotus Zhang, Jingyao Shao, Augustina Ao Liu, Lucy Jiang, Abigale Stangl, Adam Fourney, Meredith Ringel Morris, and Leah Findlater. 2022. Exploring Interactive Sound Design for Auditory Websites. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3491102.3517695

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517695

1 INTRODUCTION

Sound has unique communicative properties. Speech, song, and other human voice utterances communicate both explicit messages (e.g., words, sentences) and implicit information about the voice owner (e.g., emotion, gender, age). Sounds from natural and artificial objects can inform listeners of the sources' physical properties, location, movement, and surroundings. Different forms of sound together deliver unique, rich, and expressive auditory experiences, often in linear, spatial, and layered ways [38]. In turn, sounds are carefully designed and composed for a variety of media types (e.g., film, radio, podcasts) to create a desired atmosphere and convey accurate information.

Yet, the communicative properties of sound have rarely been explored as a means to present website content non-visually. Websites consist of content (e.g., text, media assets) and the structure of such content (e.g., hierarchy, headings), and are most commonly consumed through a carefully designed visual presentation, which we refer to as "visual websites". In contrast, opportunities to listen to an aesthetic and expressive presentation of a website's content, or "auditory websites", are nascent. Current auditory presentation is limited to simplified or serialized versions of visual websites, commonly using a *static* output style, where a single synthesized voice speaks most or all content, with only a small range of features modifiable through Speech Synthesis Markup Language (e.g., voice pitch, speed, volume [28]).

The practice of designing and composing richer auditory website experiences is largely overlooked, in contrast to the significant attention placed on the use of sound in other media forms. This gap is surprising, given the widespread adoption of voice user interfaces (VUIs) that increasingly support consumption of complex and generic content such as websites [46].

While auditory website design is in its infancy stage, the field of sound design as a whole is well-developed, with practical knowledge about people's listening preferences and sound manipulation strategies—knowledge that is in general less consulted within academic research. This paper explores aspects of professional sound design practices that can be incorporated into auditory website design, from the perspective of sound design practitioners. We

focus on investigating the following questions: (1) *How do sound design practitioners conceptualize auditory websites?* (2) *What design suggestions do sound design practitioners have for auditory website design?* (3) *What factors influence their sound design choices?*

We report on a study with 14 professionals, each of whom had at least four years of experience with sound/audio design or music production. We also required participants to have at least a basic understanding of interaction or UX design. This three-phase study included an initial interview (45 minutes), a design activity (2.5 hours) where participants created an auditory design for an existing website (either IMDB, Walmart, or New York Times), and a final interview (45 minutes). Because participants could be anchored by their current auditory interface experiences, we encouraged the use of a wide range of sound techniques that are common in interactive or more traditional media forms—such as *ambient/background sounds* (e.g., music, rainfall), *auditory icons* (e.g., short sounds associated with certain actions or functionality), *multiple simultaneous audio tracks*, *spatial audio*, *synthesized voice qualities* (e.g., perceived genders, ages, emotions, accents), *voice speed* (e.g., typical rates, slower, faster), and *volume* (e.g., louder, quieter). The interviews covered participants' design approaches for composing auditory websites, rationale for including specific sound design techniques, and overall reflections on the design activity and the future of auditory websites. Throughout, we use *sound* to mean a medium of expression, *audio* to be the experience of the sound, *sound design* as the practice of composing sound for a specific intended audio experience, and *sound design techniques* to refer to strategic manipulations of specific sound elements.

Our findings reveal five design considerations that sound design practitioners emphasized: aesthetics and emotion, user engagement, audio clarity, information dynamics, and interactivity of the auditory website. The designers experimented with a range of sound design techniques in light of these five considerations, such as manipulating voice synthesis, ambient background sound, and auditory icons (e.g., replacing spoken labels with short, memorable audio expressions, introducing variations to voices and ambient sounds based on specific sections). The designers also reflected on the challenges of applying prior sound design practices to auditory websites during the design activity (e.g., designing audio expressions for inherently visual content on websites).

Overall this paper contributes: (1) a set of design considerations that could be borrowed from traditional sound design practices to auditory website design; (2) a demonstration of novel auditory expressions proposed by professional sound designers to facilitate auditory delivery of websites; (3) sound design practitioners' challenges with applying sound design practices from traditional fields to auditory websites. These insights extend currently possible auditory website output and set the foundation for further work on mapping from sound design to interactive website presentation–including understanding the user experience of consuming more aesthetically rich auditory website compositions and developing the web design tools that would be needed to create such auditory designs.

2 RELATED WORK

Our work is informed by research on auditory (voice and/or sound) technology design and usability as well as sound design literature outside of human-computer interaction (HCI).

2.1 Designing Voice Interaction

Voice-user interfaces (VUIs) allow users to interact with a system through speech input and output. A VUI (e.g., Apple's Siri, Microsoft's Cortana) often serves as an *"assistant"* to help users *"get things done"* [59]. They are commonly used for quick information searches, listening to music, controlling smart home devices, and other small tasks [5, 12].

Despite the widespread uptake of VUIs, researchers have identified key challenges with VUI interaction. One long-standing difficulty is in the discovery and learning of new commands [26, 41, 50]. Providing the option to ask voice assistants *"What can I say?"* [26, 41], and differentiating the interaction for initial versus long-term use scenarios can help address this issue [26]. Another challenge is in supporting tasks beyond the relatively simple set listed above, such as consumption of longer content and completion of more complex tasks [16, 51, 72]. Some companies have begun to support options for voice assistants to read aloud full webpages [46], yet they adopt a static audio style, often a default synthesized voice, for all content. To allow more appropriate voice design that fits more diverse content and user preferences, researchers have begun to study people's reactions to different voice options—including voice speed, pitch, "personality", "gender", accent, and more [15, 20, 23, 66, 78]. There is also an ongoing debate on the human-likeness of voice assistants, with a number of studies indicating that human-like voices can inflate users' expectations of the system's emotional and intelligence capabilities, which can lead to frustration when juxtaposed with their actual experience [27, 44, 66].

Research on voice interaction has also largely overlooked non-speech aspects of audio experiences—such as ambient background sounds, music, auditory icons—aspects that may be useful for presenting website content and creating better audio experiences. Motivated by this possibility, this paper aims to explore the potential of richer sound design for auditory interfaces beyond what is currently available.

2.2 Designing Auditory Displays

Non-speech sound has long been used to convey digital information, such as alerting (e.g., system notifications) [35, 38], monitoring (e.g., patient data during surgery) [70], and representing data patterns analogous to visualization (i.e., sonification) [32, 74]. This practice of expressing concrete information through sound *only* is often referred as *auditory display* [38].

Over decades, auditory display research has focused on examining *techniques* to directly or indirectly map data to equivalent sound representations [32, 75] and how people *perceive* these sound representations [21, 53]. For example, researchers have experimented with using different sound dimensions, including pitch, loudness, timbre, space, and rhythm, to signify data variations [32, 70, 75]. This research suggested that each data type is best suited with specific sound dimensions (e.g., temperature is best represented through pitch), and how such sound presentation scales with data

should match listeners' mental model and perceptual capacity (e.g., the smallest pitch difference people can detect at 1kHz is under 3Hz) [75]. Early sonification work was dominated by scientific and engineering explorations that directly map abstract dimensions of sound to data values, yet such mapping may introduce difficulty for listeners without prior training to unpack the sonified information [43, 52, 63]. A range of recent work investigated new strategies that support both the aesthetics and functionality of sonification, such as by involving more pleasing sounding audio (e.g., music tones) [10, 18], utilizing everyday sounds to signal data variations (e.g., footsteps' pace [32, 75]), making use of people's innate *"cognitive schemata"* (e.g., conceptual metaphor [62]), and using sounds that provide more contexts about the sonified data (e.g., cultural value, physical property) [43].

While decades of research exists on auditory displays, they mostly focused on whether, and how, specific use of sounds can signify concrete information, mostly data-sets. Prior research had not yet considered composing and arranging these sounds for presenting complex interfaces such as websites in audio only. There is also a lack of theory for strategically using sound design to support pleasant auditory information consumption [11, 52, 75]. To close this gap, our study consulted professional sound designers on how they would use rich sounds to aesthetically and intuitively present different website content.

2.3 Audio-based Interactions for Accessibility

Screen readers allow blind and low vision users to access visual information on the screen through text-to-speech (or text-to-braille) output and fine-grained navigation control [13, 30, 72]. Screen readers theoretically offer access to any application or webpage—that is, assuming that the content has been appropriately formatted and labeled (e.g., with structural tags, alternative text). However, screen readers are expert tools that can be hard for novices to use [58]. When accessing a large amount of information, screen readers also do not support scanning [36, 49, 77] and can lead to information overload [77]. Further, issues with the underlying content design often arise, such as confusing page layout, poorly designed forms, pictures without alternative text, unlabeled PDFs, and auto-refreshing webpage components [13, 42, 45, 49].

Accordingly, screen reader users apply tactics such as increasing the speech rate, searching for information chunks, skipping unwanted or repetitive components, anchoring to a specific location on the page, and quickly listening to a page to check if it is relevant [13, 45, 71]. In terms of designing screen reader output, researchers recommended using sounds that have salient features and natural references to the represented item, to apply a small number of short, aesthetically pleasing sounds (but only when necessary), and to keep the number of synthesized voices used in representing a system small while only changing them when switching contexts [40]. In particular, past research proposed to support scanning by providing automated summary [3, 60], avoiding *"clutter"* (e.g., banners, ads), extracting important semantics [8, 60], and presenting multiple sound streams at the same time [36].

Still, the experience of accessing a website through a screen reader is generally not comparable to visually browsing a website—a visual website is often carefully designed both in terms of aesthetics and user experience, while a screen reader relies on visual website markup for content to be even accurately renderable and perceivable [30]. Our study aims to explore sound design strategies that can be applied as a base for creating more enjoyable audio-based presentations of website content. While our main focus is primarily non-screenreader users, ultimately the strategies we explore could branch out to fit different users' specific needs—a goal that requires an initial exploratory step as described in this paper as well as future evaluative studies with different sets of end users.

2.4 Sound Design Outside of HCI

Media studies has made significant contributions in theorizing sound design for artistic experience and entertainment (e.g., film, television, digital game, radio, podcast). According to the Gestalt psychology of sound, people can perceive groups of auditory objects based on whether they are temporally continuous or tend to change together, and how similar they are in pitch, loudness, timbre, and location [4]. Media sound design often utilizes these aspects of human hearing to indicate continuity within an event and closures between segments (e.g., change of music at the beginning of a new scene) [4].

In films, sound is also used to *"elicit psychological states"*, *"unify imagery"*, *"create an appearance of motion"*, and *"control attention"* [22, 25]. Three primary types of sound often appear in films: (1) *speech*, such as dialogue, monologue and off-narration, which are used directly for storytelling, (2) *music*, which helps provide an emotional atmosphere and punctuation, and (3) *sound effects*, which induce emotions and set an artificial presence [4]. These three types of sounds are also used commonly in game design, but with the additional function of providing dynamic responses to users and story context [24, 54]. Podcasts, another common media form, also provide useful sound design suggestions, especially for presenting information through audio only. For example, podcast guidelines recommend providing a short program overview, increasing engagement with theme music, introducing mental breaks, and using aesthetic tones and engaging voices [14, 29].

Looking beyond media studies, product design and urban planning also extensively consider the use of sound. For example, hybrid sports cars are often equipped with external speakers that generate engine sound feedback that can be configured by the owner for their satisfaction [67]. Urban planning also considers how soundscapes influence the perceived safety and comfort of visitors to a place [65]. Compared to media studies, these two fields seem to emphasize the hedonic value of different sounds. For example, sounds that are *"sharp"* (i.e., with high-frequency energy) or *"rough"* (i.e., with high fluctuation of sound energy) are often avoided, whereas natural ambient sounds with low frequency and temporal modulations are used to induce relaxation and positive affect [37, 48, 79]. Factors such as learned associations between sounds and emotional events, surrounding context, and listeners' status also influence whether the use of the sound is appropriate for a given context [48, 79]. For instance, familiar music is known to decrease consumers' duration

of shopping time in a department store, whereas classical music makes people spend more on luxury items [65].

Media and product design thus offer abundant sound design possibilities and recommendations that could apply—potentially with adaptation—to auditory presentation of webpage content. We explore this potential by inviting professionals with sound design experience to envision how webpages can be presented through audio using a richer set of sounds than is typically used in VUIs and screen reader output.

3 METHOD

To understand how professional designers envision rich sound for audio-based delivery of webpages, we conducted a three-part study that included a design activity to create an auditory version of an existing webpage, together with initial and final interviews.

3.1 Participants

We recruited professionals who had "experience with audio/sound design or music production" and met the following inclusion criteria: experience with user experience (UX) design or web design (i.e., some level of familiarity with interactive design concepts), experience with audio editing tools (to ensure that they could complete the design task within the assigned period), and experience with using smart speakers (e.g., Amazon Echo, Google Home). We utilized two freelancing platforms, Upwork [19] and Fiverr [33], for our recruitment.

We initially enrolled 15 experienced designers, one of whom dropped out midway due to a personal reason. The remaining 14 designers' self-reported experience with sound design and audio engineering range from 4 to 20 years (*Mean*=12, *SD*=4.42). More specifically, 13 participants' past experience was in audio, music and sound production, while P3 worked in the *"podcasting"* (P3) industry for 11 years with a focus on *"sound mixing"* (P3) (full details of participants' sound design experience are included in Supplementary Materials). For audio editing tools, participants commonly use Pro Tools [68] (N=10), Logic Pro [6] (N=6), and Garage Band [7] (N=5).

In terms of experience with interaction and UX design, seven participants (P1,P2, P3, P5, P6, P7, and P14) self-reported having less than one year of experience, four participants (P4, P8, P11, P12) had 1-5 years of experience, and three participants (P9, P10, and P13) had 5-10 years of experience. All but P2 used voice assistants (either on smart speakers or smartphones), but the frequency of use varied. P4, P5, P7, P8, P9, P13, and P14 used voice assistants at least once every day, whereas the remaining participants used voice assistants only sporadically.

3.2 Procedure

The study procedure was administered remotely and took four hours, including: an introduction and interview (45 minutes), an audio website design activity (2.5 hours), and a final interview (45 minutes). The final interview was scheduled 24-48 hours after the initial interview, depending on each participant's availability. The second and third author collaboratively conducted the interviews. The study was approved by our university's Institutional Review Board. All participants provided informed consent and were compensated $100 for their time. Study materials can be found in Supplementary Materials. Below we detail our study procedures at each stage.

3.2.1 Introduction and Initial Interview. The initial interview prepared participants for the design activity by introducing the idea of richly designed auditory websites and by describing their specific design task.

We first prompted users to imagine future audio-based websites that:

- *"Provide audio/voice information that is equivalent to a visual website, such as conveying aesthetics, brand identity, emotion, and hierarchy. That is, the audio version is not a simplified version of a visual website."*
- *"Go beyond the current approach of using a single voice and speech rate for the whole page. For example, consider how non-speech audio may be incorporated or different voices may be used for different webpages, types of content, etc."*

To further concretize this idea, we described an *envisioned voice-based browser* (Table 1) that would allow users to interact with these auditory webpages. This envisioned browser would run on a smart speaker or other device, have the same basic functions as a traditional web browser but with speech for input and audio for output, present webpage content using synthesized speech and/or other sounds, and allow the user to pause or jump around to different sections of the page. To show how the mechanics of this browser might work, we then played a 75-second audio clip of a voice-based browsing scenario (Table 2). This sample clip only used a single *default* synthesized voice to present all information (i.e., similar to default commercial voice assistant sound), which we contrasted to the task we were asking of participants: *"As you heard, the single synthesized voice and speech rate used in this example did not have the same aesthetic richness as a visual website would have had. This is the problem we want you to address [in designing a new auditory website]."* We asked about the participant's initial reaction to the idea of a voice-based web browser, and how they might approach the design of auditory webpages to be consumed in this way.

We then introduced participants to a specific webpage to design. Because different websites have vastly different branding styles and complexity, we selected three contrasting webpages and randomly assigned each one to roughly one third of the participants: the homepage of the *New York Times* (*NYT*), the page for the movie *Titanic* on *IMDB*.com, and *Walmart*.com's product page for the TI-84 graphing calculator. We asked participants to describe the overall visual style and branding of their assigned website, and how they might convey them through an auditory version of the page.

Finally, to help participants think broadly about audio design possibilities and reduce the risk they would be anchored by experiences with current voice interaction, we presented a range of sound design techniques to consider using: *synthesized voice qualities* (e.g., pitch, emotions, accents), *voice speed*, *volume*, *ambient/background sounds*, *auditory icons* (i.e., short sounds associated with certain actions or functionality), *multiple simultaneous audio tracks* (e.g., foreground and background), and *spatial/stereo audio*. We provided brief audio examples of these ideas to participants and asked them to comment on which seemed particularly useful or not useful. Participants further described any other audio element ideas they had.

Table 1: Basic voice browser commands that we presented to participants before the design activity to convey the mechanics of how the envisioned voice-based web browsing might work.

Basic Voice Commands	Function
"Load [title or URL]"	Load a page
"Get overview"	Read out the overall structure of the website with major headings, along with a short text description of the page if available
"Pause"/"Continue"	Pause or continue reading the content on the page
"Jump to [item]"	Jump to a named item in header, main content, or footer and continue reading
"Find [X]"	Find content X within the page
"Select [item]"	Select the named item
"Open"/"Close"	Open/close menus, sections, etc. to get more or less detail

Table 2: An example scenario of using voice-based web to browse a specific university website. We emphasized that the scenario of use audio clip included only a basic synthesized voice to read the content aloud, but that participants would be encouraged to employ aesthetically richer voice and sound design.

User's Input	Browser's Output
"Load [anonymous URL]"	*"Loading [URL] [University Name], Search, Quick Links"*
"Get overview"	*"Navigation Bar, Featured Stories, News and Events, Fast Facts, Connect"*
"Open Navigation Bar"	*"Search, About, Academics, Apply, News and Events, Research, Campuses, Give"*
"Open Academics"	*"Opening Academics. About [this university], Colleges and Schools, What are you driven to discover? A life-saving cure? An entirely new art form? A solution for greener technologies?"*
"Go back to home page"	*"Returning to [URL]. [University name]. Search"*
"Jump to news"	*"News and Events. July 1, 2019. How you and your friends can play a video game together using only our minds. Read more. June 27, 2019. Astrobiology outreach. [University name's] mobile planetarium lands at space conference. Read more"*

This session concluded with specific instructions for the design activity, described next.

3.3 Design Activity

Participants were given 2.5 hours to create an audio clip mockup (~2 minutes long) to convey how they envisioned their assigned webpage (i.e., NYT, IMDB, or Walmart) should sound when consumed through audio. We prompted participants to focus on how the content should *sound* rather than on improving how users would *interact* with that content (e.g., coming up with new voice commands).

As a basis for the mockup, we provided content from the visual version of the website. Rather than providing the original full visual website, we took screenshots of each section of content (e.g., news stories, movie synopsis, product information) and presented these linearly in a Google Doc (Figure 1). The goal was to encourage participants to create a fundamentally auditory website design, rather than focusing on the original visual design of the website.

To structure the 2.5-hour design process, we invited participants to first spend 30 minutes envisioning how each section of the webpage would sound and record their design ideas in the aforementioned Google Doc. Participants then mocked up a 90-120 second long audio clip to demonstrate what it would sound like for a user to visit the audio webpage that they envisioned. They could use any audio editing tool they wished, and could focus on mocking up only a subset of the sections as long as at least one full section was included.

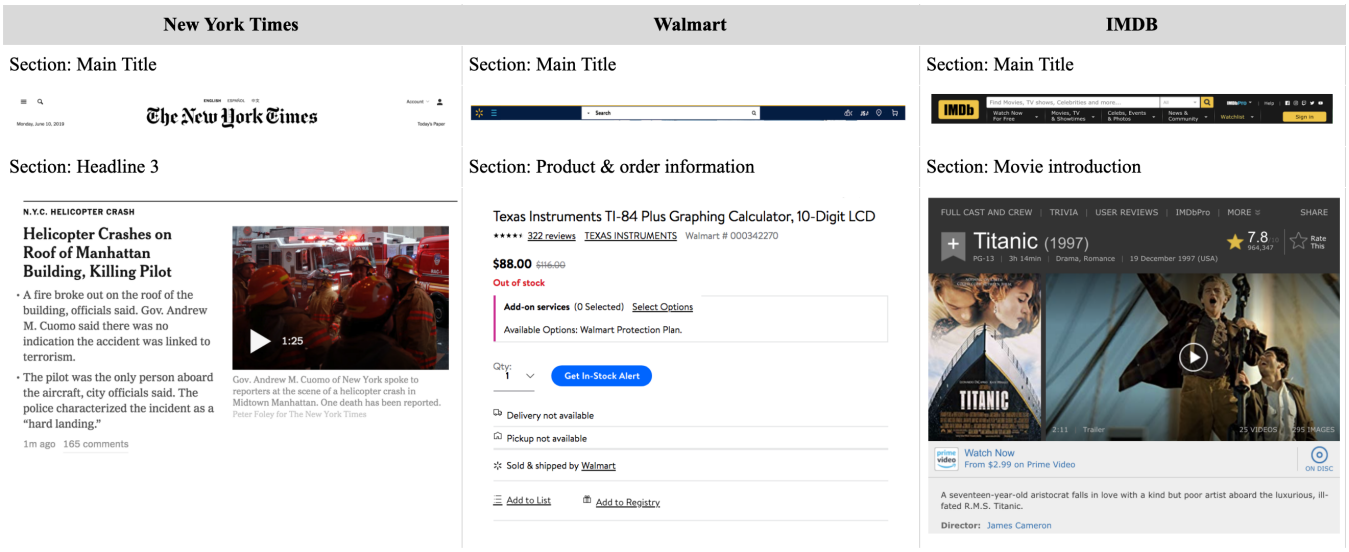
Throughout our instructions, we emphasized that the task was to create a "mockup" that represents a participant's design ideas as closely as possible, but that did not have to be high-fidelity. To support participants' mockup creation, we provided a set of resources for outputting text with different synthesized voices (e.g., IBM Speech to Text [39], Readaloud [61]) and for finding Creative Commons licensed sounds (e.g., Freesound [34], Zapsplat [76]). Participants submitted the finished audio clip to the research team via the freelancing platform before the final interview.

3.3.1 Final Interview. The final interview (45 minutes) began with the participant's professional background and experience with voice interaction. We then asked them to describe their design and reflect on: whether they had thought about the branding for the site, how the audio clip might be different if they had more time, and why they did or did not use a variety of audio elements, including the ones we had introduced in the initial interview. The interview closed with more general questions about the design of the auditory webpages, such as how design choices would change for other websites similar to the one the participant had focused on, how those choices might differ for other classes of websites, and thoughts on the voice interaction and commands (as opposed to the primary focus of the study: content).

3.4 Data and Analysis

The interviews were transcribed by the second and third authors and a professional transcription service. As design theory for audio-based website consumption is sparse, our interview and design

Figure 1: Participants designed their auditory webpage based on content extracted from visual versions of those webpages. To reduce the influence of the original visual layout, this content was presented as separate screenshots for each section of the original page, laid out linearly in a document. Example sections are shown here from the New York Times, Walmart, and IMDB websites.



activities are exploratory. Therefore, we adopted an inductive thematic analysis approach as outlined by Braun and Clark [17]. In the first phase of analysis, the first author read and re-read the interview transcripts to identify an initial set of codes, with inputs from the second and third author (who had conducted the interviews). The research team then collaboratively developed an initial codebook to guide the coding activity. The first author individually coded all of the transcripts, with the fourth author selecting (based on a random number generator) half of the coded transcripts to review whether the coding and overarching themes accurately depict the data. This review resulted in minor conflicts in the application of three sub codes, which were resolved through meetings between the first and forth authors. The first author then defined each theme, adjusted the coding of the rest transcripts, and organized them into the findings. The final codebook includes six overall themes: sound design techniques, technique application, design considerations, site-specific considerations, auditory website design conceptualization, and challenges—The full codebook is included in Supplementary Materials.

To analyze participants' mockups, the first and fourth authors independently identified sound design features of each mockup, then collaboratively reviewed the identified design features, focusing on: (1) usage of specific sound design techniques, (2) the overall design approach, and (3) the content of each prototype.

4 FINDINGS

Our data collection focused on understanding how sound design practitioners conceptualized auditory webpage design. We first provide an overview of how participants designed their auditory website mockups, then describe sound design choices they made for supporting five commonly prioritized design considerations. Last, we report on how sound design practitioners in our study reflected on their experience and challenges during the design activity.

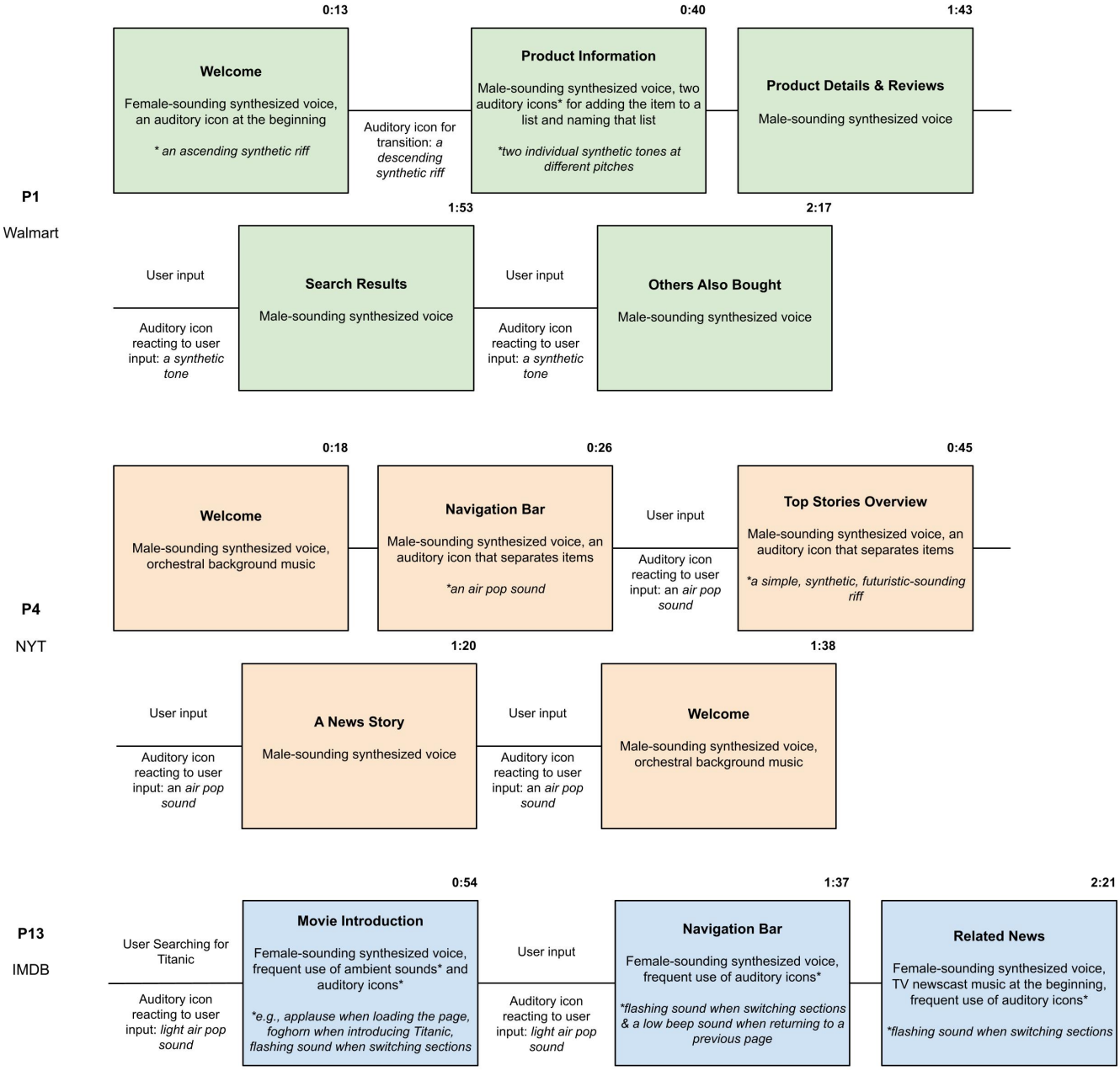
4.1 Overview of Sound Design Practitioners' Auditory Website Mockups

This section provides a quick summary of participants' mockups, focusing on their content composition, usages of sound design techniques, and design foci. Of note, the design mockups were intended to be more about process than outcome—low-fidelity artifacts limited by time and available tools yet allowed participants to deeply engage with their sound design ideas. Still, they provide context for later sections on participants' design practices and rationale.

4.1.1 Composition of auditory website mockups. Participants' mockups contained between 1 to 13 sections (*Median*=4) and ranged from 90 to 924 seconds in length (*Median*=120.2s—P8 chose to mockup all provided sections instead of only a few as we suggested, resulting in a mockup that was 924s in length). The majority of participants chose to work with the most central content of the web page they were assigned to, such as movie information for the IMDB Titanic page, headline news for the NYT home page, and product descriptions for the Walmart calculator page. Figure 2 shows three example mockups, while Supplementary Materials includes all 14.[1] The mockups featured voice narrations of website content (N = 14), sound effects (N = 13), and conversational interactions (N = 6). Additionally, nine participants included a site overview in their mockup, and nine participants also created a short welcome clip.

[1]Some audio files of the mockups contain participants' own voices which cannot be disclosed based on our consent with participants.

Figure 2: Visualizations of example auditory website design mockups that participants created for the New York Times, Walmart, and IMDB websites. Note that we used this time-constrained design activity as a methodological technique to engage designers more deeply in the design process rather than as a means of creating a complex, refined prototype; the sound techniques used in the mockups may thus skew toward easier-to-implement techniques. Consecutive sections with shared sound design settings were combined into one box for brevity. End timings for each section are shown at the top right of the box in minutes and seconds.



For example, as shown in Figure 2, P4's mockup was 98 seconds long, beginning with a male-sounding synthesized voice welcoming users to the New York Times website and classical orchestral music playing at the background, followed by a quick, simple overview of navigation options, a read-aloud of top news story headings and content from one specific news article, then returned to the welcome section, as the user commands. P13's IMDB mockup instead directly began with an introduction to the movie *Titanic*.

4.1.2 Usage of sound design techniques. Table 3 summarizes the sound design techniques participants planned to use and actually used in their mockups. Overall, all seven optional sound design techniques we provided were used by at least one participant, while four participants additionally wanted to add audio effects (e.g., reverb, delay, chorus) as part of their sound design techniques. Of these eight techniques, four were used more in participants' mockups than the others (Table 3): *auditory icons*, *ambient sounds and music*, *multiple audio tracks*, and *change of synthesized voice qualities*. For the other sound design techniques, some participants were particularly concerned about potential confusions and inconvenience brought by them ($N = 6$ for spatial audio, $N = 4$ for unstable volume, $N = 3$ for speed)—*"I think it would be a huge inconvenience to have to change or turn their volume knob for different sections"* (P8). However, for the majority of participants, sound volume, speed, and spatial quality are still useful properties to manipulate, which they planned to, but were unable to incorporate, due to: (1) limited time and available tools during the design activity; (2) expectations that their prototypes would be low-fidelity—*"I didn't mess around with [the] speed of it because it's almost like a placeholder for what the real design would be"* (P7). Although time-constrained, the design activity still provided participants opportunities to engage deeply with auditory website design and come up with constructive ideas. Therefore, the focus of this paper's analysis is more on participants' design rationales and perspectives than the exact sound technique usage.

4.1.3 Design foci. Participants' use of conversational interactions was more sparse compared to mainstream voice assistant interfaces, typically only for providing navigation support and quick response to a user question. Instead, participants reported relying on *"podcasts"* (P9, P2, P7), *"radio"* (P2, P13), and *"storytelling"* (P4, P5, P10) as metaphors to conceptualize their envisioned auditory website experience, generally reflecting their experience with these non-interactive sound design fields.

We observed a set of design considerations that sound design practitioners in our study commonly mentioned and prioritized, including aesthetics and emotion ($N = 14$), user engagement($N = 14$), audio clarity ($N = 14$), information dynamics ($N = 13$) and interactivity ($N = 14$). Their choices for audio styles, voices, and sound design techniques were formed based on these considerations, which varied for different websites and content, as P12 indicated: *"the target audience is different, the way people use it (the website) is different. So for sound design, it's usually on a per-site basis, based on what it's being used for. So what could be considered too many sound effects on the New York Times website, on a social media website is just considered normal"* (P12). In the following sections, we detail how they used different sound designs to facilitate each specific considerations across the three different sites.

4.2 Sound Design Choices for Conveying Aesthetics and Emotion

As visual website designers use colors, pictures, and icons to present content topics, aesthetics, and emotion, our participants attempted to create equally rich experiences in audio. They used a range of techniques to shape the style and feeling of a website:

First, ambient sounds and auditory icons were often used for delivering specific effects, such as *"door slamming/locking"* or *"witch cackle"* (P3) for spookiness and *"hustle and bustle from the street"* (P10) to represent New York city. Participants often layered ambient sounds or auditory icons with specific voice options to communicate the content topic of a page, such as *"field recording of crickets, cicadas, and wind blowing through trees"* (P10) to represent nature, *"the sounds of light chatter and dishes"* (P3) to represent restaurant, and *"crowd cheering sound"* (P1) to represent a basketball game. Five participants proposed to play specific products' sounds as users browse an audio shopping site, as P7 did for the calculator page on Walmart: *"Maybe the computer, electronic devices, or smart devices have some synthesized computer sounds, where on some level it conveys that this thing is a calculator."* P7 and P10 further mentioned conveying the environmental characteristics of a place by recording or constructing its soundscape: *"I think there's a lot of interesting things that can be done with skilled recording to give you an idea what it sounds like in Yellowstone for example"* (P7). Music also naturally conveys emotional information, as P10 suggested: *"If you want someone to feel happy, your chord progression should go upwards, and [if] you want someone to feel sad, they'll go downwards."*

Designers also felt that the personality, gender, tone, and accent of narrating voices could contribute to the overall feeling of the site. For instance, IMDB designers P10 and P5 used an *"older gentleman"* (P5) sounding voice to create a serious storytelling atmosphere. P7 also suggested to use a *"British voice"* (P7) to deliver content that is associated with British culture, such as a *"British Museum"* (P7) webpage. Participants considered how the choice of voice influences users' own emotions too. For example, P1 pointed out that the tone of a voice can be critical for delivering emotion-eliciting messages: *"For sales, you want someone enthusiastic. For a complaint, you want someone apologetic."* P5 and P10 also proposed to use a trustworthy-sounding voice, such as the default Alexa or Siri voice, to deliver important information that intends to be trustful.

The overall audio and voice styling of participants' mockups varied greatly based on specific web pages' aesthetics and branding. At the onset of the study, participants described their impressions of NYT as *"straightforward"* (P2, P4, P9), *"clean"* (P2, P12), *"busy"* (P2, P9), *"classic"* (P2, P9), and *"trustworthy"* (P2). To match the auditory website's style with NYT's branding, all participants chose *classical* background music and auditory icons (e.g., piano riff, orchestral music). Similarly, Walmart designers chose a minimalistic sound design to fit the *"simple"* (all 5 participants) and *"organized"* (all but P14) styling of the Walmart brand. At the same time, they used more *"exciting"* (P7), *"enthusiastic"* (P1), *"high quality"* (P5), and *"non-robotic"* (P7) voice tones to facilitate a pleasant customer shopping experience. For IMDB, participants also wanted to keep the parts of the page irrelevant to movies (e.g., general navigation menus) simpler and *"non-branded"* (P8) to fit the website's functional branding. However, participants consistently discussed the

Table 3: The sound techniques participants included in their design plans (i.e., Google Doc template) during the design activity (marked in "x"), alongside the techniques that actually appeared in their mockups (marked in "o"); the techniques are ordered by how many participants planned to include each one in their mockups. As already noted, the sound techniques used in the mockups may skew toward easier-to-implement techniques, thus it is important to consider both the plans and the mockups. Among the three sites, NYT had less varied voice qualities than Walmart and IMDB, whereas Walmart designers used ambient sound and multiple audio tracks relatively less. The techniques here include the seven that we prompted participants with in the design task, plus one emergent technique (Audio effects).

ID	Site	Auditory icons	Multiple audio tracks	Ambient sound	Change of voice qualities	Change of voice speed	Change of voice volume	Spatial audio	Audio effects
3	IMDB		x o	x o	x o	x	x	x	
5	IMDB	x o	x o	x o	x o	x	x		
8	IMDB		x o	x o	x o	x o			x o
10	IMDB	x o	x o	x o		x	x		x o
13	IMDB	x o	x o	x o				x o	
1	Walmart	x o			x o	x	x		
6	Walmart	x			x o	x o		x	
7	Walmart	x o	x o	x o	x	x	x o	x o	
11	Walmart	x o	x o	x	x o				
14	Walmart	x o			x o	x	x o	x	x o
2	NYT	x o	x o	x o	x o	x	x o		
4	NYT	x o	x o	x o		x	x	x	
9	NYT	x o	x o	x o	x o				x o
12	NYT	x o	x o	x o	x	x	x o	x	
Plan total	-	12/14	11/14	11/14	11/14	11/14	9/14	7/14	4/14
Mockup total	-	11/14	11/14	10/14	9/14	2/14	4/14	2/14	4/14

need for expressiveness when conveying movie-related content (e.g., storyline, production quality). They tended to add expressive, artistic elements, such as playing a variety of representative movie sounds, including *"ship horn"* (P10), *"ocean sounds"* (P5), and the Titanic's theme song *"My heart will go on"* (P3). There was also a preference of using voices that sound close to the age, gender, and accent of the main characters or that have a tone appropriate to the story. For example, both P5 and P10 wanted to use a serious voice for Titanic given the seriousness of the movie.

While participants only worked with one website in this design activity, they shared that for other websites of the same category, they would change stylistic choices based on these sites' branding too. For example, NYT designers suggested that they would make the audio styling more *"lighthearted"* (P12), *"country-sounding"* (P11), *"bombastic, strong, and fierce"* (P4) for Fox News, and more *"dry and boring"* (P12) for CNN. As another example, P3 and P5, who had focused on IMDB, felt that they would make the audio more fun and involve more community elements for the competitor site Rotten Tomatoes, as it focuses more on opinions compared to IMDB.

4.3 Sound Design Choices for Engaging Users

Another design consideration raised by our participants is how to engage users, or listeners. All 14 participants were concerned that users may become disengaged with prolonged audio interaction. This concern arose especially with synthesized voices, such as *"it would be easy to lose focus or to stop paying attention to what the voice is saying"* (P7). They commented on how the lack of visual stimulation may make focusing harder for users who are not used to pure audio information consumption, *"especially in the age that we live in, where people are just so visual that people would rather watch a movie than read a book"* (P5).

Our participants attempted to prevent users from disengaging by introducing variations in their designs, an approach described as *"pattern disruption"* by P4. One example of pattern disruption is switching off voices for different sections: *"just to make it seem like there was more than just one robot voice reading the entire script"* (P9). The variation can be in the voice personality, gender, or other qualities: *"Maybe it's a team of voices or different personalities that talk about those different things that I think that would be very engaging"* (P4). Participants also suggested involving audio elements beyond speech, such as *"music to just keep the energy up"* (P12), as presented in the previous section.

At the same time, our participants were concerned that overly stimulating, unpleasant, or irrelevant information could exacerbate users' lack of engagement. Thirteen of them shared that they would make the audio comfortable to listen to with soothing background music, natural sounding voices, smooth transitions, balanced volume, and consistent speed: *"For a comfortable setting, I would imagine that would be at pretty much a normal speaking rate of voice - not whispering, not yelling, but just a normal tone of voice as I'm using right now"* (P6). Five participants further suggested that involving real human voices and natural dialogues would elevate the experience: *"just hearing like a computer read off something, it's just not as a satisfying or you don't really get the human interaction as much"* (P9).

Among the three websites, NYT and IMDB designers paid particular attention to users' listening comfort and engagement. Participants pointed out that NYT's content is mostly longer text and thus tried to retain listeners' attention with auditory icons and

varied voices that indicate transitions to new sections, as well as more natural-sounding and pleasant voice choices: *"if resources were of no concern, [the sound] would be completely tailored to the audio experience of consuming news"* (P7). For IMDB, participants wanted to create a *comfortable, engaging* experience appropriate to browsing *"during their (users') leisure time"* (P8). They therefore kept less important information (i.e., content irrelevant to movies) in a simpler presentation format to focus users' attention.

4.4 Sound Design Choices for Ensuring Clarity

All participants considered audio clarity to be critical, as P2 suggested: *"The intelligible quality of a voice, I think that's probably the most important. I don't want to listen to a menu and not be able to understand it"* (P2).

Participants focused particularly on synthesized voice qualities and soundtrack arrangement to improve clarity. Many suggested using a slow but consistent speech rate for clarity: *"I'd use a slower voice for information that you would want to make sure that you could take in well"* (P14). P10 and P12 further suggested adding breaks and gaps between groups of information to help understanding, while P6 wanted to provide users the option to re-listen to previously played content. The majority of participants also wanted to avoid too many soundtracks and distracting background music, as they could cause sensory overload and in turn diminish clarity of important speech: *"If we have multiple audio tracks going at once, I feel like it could be very, very confusing for the listener"* (P1).

To improve audio clarity, all but one participant considered allowing users to customize the synthesized speech—especially speed and volume but also possibly other qualities (e.g., gender, accent, age). They commented that individuals have different listening capabilities and preferences that can critically affect their user experience, reflecting past research findings [15, 20, 78]. Cognitive and sensory abilities as well as cultural-language background were all factors mentioned by our participants as influencing users' preferences for voice speed and volume: *"People in New York tend to talk quicker than people in Texas. So if you could have control on voice speed, that would be great"* (P2). Participants also commented that the listening environment (e.g., noisy, P2) and audio equipment (e.g., headphone vs. external speaker, P12) could impact audio clarity. For this particular instance, P2 recommended using a higher-pitched voice as it *"cuts through"* better: *"I know if I'm driving in a car, a lot of low end male voices don't cut through as well, because of the road noise"* (P2).

While audio clarity was important to all three websites, NYT designers were particularly careful with choosing a clear, understandable voice to ensure the clarity and accuracy of news reading—*"somebody who was easy to understand, spoke clearly and pleasantly, and seemed like [a] fit [to] the brand of the New York Times"* (P4). All four participants working with NYT kept their background music light and simple. They all mentioned not wanting to use too many soundtracks and needing to balance each track's volume: *"Three elements is probably the max that I would use"* (P2). For certain critical information on Walmart, such as payment functions, our participants were also particularly concerned about the potential for interaction errors and thus kept clear voices and minimal distracting soundtracks as priorities.

4.5 Sound Design Choices for Indicating Information Dynamics

In text-based visual media, important information may be bolded, italicized, presented in large, eye-catching fonts, put into a central location, or emphasized by surrounding white space. This visual formatting captures the dynamics of the information. Correspondingly, all but one participant in our study considered representing the importance level of different information in audio, as P14 said: *"Making the audio dynamic to fit the dynamics of the text."*

Strategies for presenting these dynamics included switching ambient background sounds and music on and off, layering distinctive auditory icons or sounds, introducing different voices for important content, as well as changing volume, speed, and spatial quality of the voices. P10, for example, suggested not to use any ambient background sound when presenting less relevant information. P14 considered adding a reverb effect for content with a larger font. A few participants proposed to have a specific synthesized voice to read out important information, such as a *"male voice for the titles and all of the bolded font on the page"* (P14) or a *"robotic"* (P3, P6) voice for facts about movies: *"I was looking for something that sounded, I wouldn't say robotic, but sounded like it was being read and let people know—Hey, we're delivering (factual) information"* (P3). Eight participants also mentioned turning up voice volume for important information such as section headers and product descriptions while lowering the volume for peripheral content such as sponsored ads. Similarly, 12 participants considered adjusting speech rate to reflect the importance of spoken content, with faster speech for less important content. A few participants also mentioned that spatial audio could help to differentiate sections of the website, such as: *"'Oh, now I know I'm listening to menu stuff because it's coming from the left'"* (P5), and that audio effects such as reverb could be used to emphasize important information: *"I used a little bit of reverb on the male voice just to make it obvious that it was a headline kind of font"* (P14).

4.6 Sound Design Choices for Facilitating Interactivity

While performing sound design for interactive interfaces is a new task for all participants, they took efforts to consider how to facilitate interactivity of an auditory website. Most commonly, they considered three aspects of website interactivity: status of the interface ($N = 14$); navigability ($N = 13$); interaction efficiency ($N = 12$).

4.6.1 Status of the interface. Participants considered various ways to inform listeners about the status of the interface (i.e., visibility [55]), such as actions that users could enact on specific website content and the website's reaction to such actions. For example, P6 and P8 proposed to use voice variations and auditory icons to indicate hyperlinks: *"adding an audio effect to the voice to indicate that this is a hyperlink, this is something you can proceed to... maybe add a little delay or chorus, just a tiny bit for people to know 'oh this phrase actually represents a hyperlink'"* (P8). P7 also used a repeated chanting sound to indicate that a search is being processed. To help users better distinguish among different statuses, participants tried to come up with sounds that are distinct. For example, to indicate

that the user has "clicked" on a paid NYT article, P9 specifically wanted to use a *"cash register"* (P9) sound. Walmart designer P7 further suggested mapping unique auditory icons to each product.

4.6.2 Navigability. The majority of participants envisioned it being difficult to perceive the overall structure and available navigation options in audio: *"overview, news, updates, get to know us, contact us—they have to hold all of those options in their short term memory just to be able to make one selection"* (P5). To relieve the burden of navigation, participants recommended reducing available options and *"to go very simple with what the menus were doing so that the focus was on the content"* (P5). Some of them used conversations to guide users' navigation. For example, six participants included dialogues when prompting users to move to a new section, such as *"Would you like to see more?"* (P7's prototype). Many participants also considered how to use sound design to make it more intuitive to users *where* they are on the website, as suggested in 4.2. P1, in particular, suggested layering multiple soundtracks (e.g., different instruments) to: *"let a person know how deep they were into the website—as you (the user) backed up, maybe there would be less music."*

4.6.3 Interaction Efficiency. Reflecting past work [36, 49, 77], the majority of participants were concerned that engaging with information in audio could be time-consuming, especially when they need to quickly scan a website, as P7 shared: *"Because you can only take in so much in real time over audio, versus visually you can take in a lot more information faster."* To speed up the interaction, many participants suggested reordering information, so that key sections would be heard immediately without having to *"wade through a bunch of extraneous stuff"* (P4). Many participants also proposed filtering out unimportant or tedious information, such as ' *'see more"* (P10), *"the bottom half of the page"* (P6), and *"all 34 reviews"* (P7), to result in *"every single thing having a purpose"* (P10). Similarly, some proposed to speed up less important information to save time. Many also chose to play a short, representative auditory icon, ambient sound, or music clip rather than speech to quickly signal specific feedback or page sections— *"because it can get tedious to just listen to every instruction [in speech]. But instead using sound that people can really relate to what it means can connect it better with the function"* (P11). Participants commonly used existing associations between sounds and concepts, such as *"the CNN background music"* and *"Anderson Cooper"* voice proposed by P3 as *"something that people feel a connection with"* for the news channel, CNN. Further, P1, P5, P7, and P12 all felt that when looking for specific information, being able to directly ask the voice assistant is easier and more efficient than needing to listen through the website: *"You know, 'Siri, tell me who was the lead actress in the Titanic.' I don't think they're gonna want to access the website and listen to five minutes worth of data to get that"* (P5). Finally, P5, P7, and P10 all proposed to adapt the interface based on how familiar a user is with the site, such as: *"The first time they go to the 'electronics' page, the audio branding starts for the 'electronics', and a voice comes on and says, 'You've reached the electronics page ...' If I keep going back to the 'electronics' page to shop for electronics, I don't want to hear that voice every time I go back"* (P7).

Among the three websites, Walmart designers especially focused on efficiency as the top priority, as they envisioned Walmart users' goal to be efficient shopping, as P11 shared: *"I just wanted to make it straight to the point for people who need to get this calculator"* (P11). They focused mostly on portraying the product features (e.g., brand, price, functions) and excluded unnecessary information such as *"all 34 customer reviews"* (P7) and *"about us"* (P1) toward a minimalistic design. Simple auditory icons (e.g., a short riff or tune, clicking sound, air popping sound, bell ring) were often used to replace spoken words as feedback on user actions and to signal transition between sections.

4.7 Challenges with Sound Design on Auditory Websites

In reflecting their applications of sound design practices on interactive websites, designers mentioned several main challenges, including *bias from previous engagement with visual websites*, *challenges with inherently visual content*, and *difficulties with sound design related to the abundance of website components*.

4.7.1 Bias from previous engagement with visual websites. Four participants confessed that it was difficult to imagine a fundamentally auditory version of a website that they have previously visually engaged with. P9, for example, found it challenging to assess his prototype: *"A better assessment could be made if someone were to listen to it without looking or knowing any visual representation of the webpage"* (P9). Many designers felt biased by the visual design of the website. P9 always envisioned a website to have a visual navigation bar, whereas P1 kept going back to the idea of representing *"how deep a user is into the website's structure"* (P1). More than half of the participants considered the existing website's visual style (e.g., color use, font size) when performing the sound design, while P6 even tried to *"go from top to bottom and left to right to give the user a sense of space how a page would look if they were able to see it"* (P6).

4.7.2 Challenges with inherently visual content. Participants criticized the website materials we provided for the sound design activity for being inherently visual-based rather than audio-based, as P5 noted: *"So this [visual] website is based on conveying concrete information that's divided up into sections or headings, but I think that an audio based website has to be something that doesn't feel so constrained by space"* (P5). Certain visually-bound content can be particularly difficult to convey through audio, such as a *"site overview"* (P9) and *"a map"* (P5). P7 further commented: *"the interactivity and the user experience would just be very different, because you can only take so much in real time over audio, versus visually you can take in a lot more information faster."*

4.7.3 Abundance of website components introduces difficulty to sound design. Almost half of our participants brought up the difficulty of choosing auditory icons and ambient sounds. This challenge could aggravate on websites with many distinct sections, as P2 spoke: *"I think it was the abundance of content on the New York Times site that made it feel like there were too many auditory icons"*, especially when the auditory icons need to be distinctive enough for users to recognize but also fit the overall styling of the website. Some participants tried to incorporate existing commercial music or sound clips of a company as a solution, to *"keep the branding consistent"* (P9).

5 DISCUSSION

Through a series of interviews and design activities, we explored how experienced sound design practitioners approach the creation of auditory websites. Our findings reveal a set of design considerations they prioritized for auditory website design as well as creative ways to manipulate sounds for each consideration. These design considerations and techniques draw insights from traditional sound design practices to inform the design of emerging auditory websites. Here, we discuss how our participants' perspectives align with and differ from design foci of auditory interfaces in HCI, and present promising research directions for future auditory websites.

5.1 Auditory Interface Design in HCI and Sound Design for Non-interactive Media

Existing auditory interface research in HCI focuses primarily on three types of applications: VUIs, auditory displays, and basic audio-based access to *natively visual* digital content (i.e., a screen reader speaking a webpage). Each application area has emphasized different aspects of interaction—VUI research centers around usability of voice commands (e.g., [26, 41, 50]), synthesized voice qualities (e.g., [15, 20, 23, 66, 78]), as well as how VUIs influence varied aspects of users' life (e.g., [5, 12, 59]); auditory display research mainly comprises sonification studies that aim to intuitively represent values of a dataset through sound, on par with visualization (e.g., [32, 63, 74]), and experimentation around non-speech sounds for signifying certain events or object features (e.g., [35, 38, 70]); screen reader research has focused on such tools' basic functionality, centering the experience of blind and low vision users (e.g., [49, 72, 77]). Together, this collective body of research has a focus on usability, functionality, and end user experiences.

In contrast to those traditional HCI focus areas, professional sound design practitioners in our study emphasized on a number of considerations that are often left out in existing auditory interfaces. In particular, aesthetics, emotion, and listener engagement were prioritized much more by our designers compared to prior HCI auditory interface research. Aesthetics and engagement are both important factors influencing the adoption and experience of technological products [9, 69]. Aesthetics was less explored in the early days of auditory displays, but has gained increasing appreciation (e.g., [63]). Recent sonification researchers argue that aesthetics of audio display not only adds marginal value (e.g., reducing annoyance), but also support listeners' meaning-making of sonified data [63]. This line of research thus calls for attention on designing more intuitive and aesthetic auditory displays [63, 64]. Our study extends existing progress toward this goal, which has focused mostly on data sonifications, by contributing to intuitive and aesthetic auditory presentations of *websites*, drawing professional sound designers' expertise.

Designers in our study proposed many creative, exploratory sound design choices to support both aesthetics (including style, feeling, and listener engagement) and more functional aspects of auditory websites at the same time—including aspects that prior HCI auditory interface research has deemed important too: navigability, efficiency, and information dynamics [26, 36, 41, 49, 50, 77]. For example, background music and ambient sounds representative of different websites' content could facilitate emotion-eliciting presentations and immersive experiences, and may also help to indicate the type of content being played (e.g., playing a movie's soundtrack while on an auditory IMDB webpage would indicates what movie the page focuses on), contributing to navigability. Further, variations in narration voices, soundtrack arrangement, ambient background sounds, and music could make the audio more engaging, and at the same time also help to convey information dynamics and website status, when such variation is based on the structure of a website. Many of these strategies are inspired by designers' past experience from media production fields. For example, the use of music and ambient sounds to prime audiences and create closures across sections reflects film production practices [4], whereas the introduction of music and mental breaks is often used in podcasts [14, 29]. Together these sound design practices provide new insights and possibilities for future auditory website design research to reference.

New concerns also emerged through our investigation into sound design for auditory, interactive websites–concerns that will only be fully understood with future work that examines end user experience. As noted by many of our designers, complicated sound effects and soundtrack arrangement may introduce confusion and inefficiency for end users. Clarity and accuracy will likely need to be prioritized for critical information (e.g., that on a check-out page of a shopping site, key news items), while simple, quick auditory presentations may be needed in some scenarios (e.g., when a user is in a rush). Moreover, users who rely on screen readers and auditory interfaces as their main means of technology access usually prefer a faster speech rate [15] and may have other preferences that differ from those of sighted users who can fall back on a visual version of a website if desired. Therefore, designers should not only pay careful attention to what requirements each website's content and main usage imply, but also support users' agency and control over auditory websites' volume, speed, background music, and number of sound tracks to ensure basic usability. Further work is needed to understand these end user perspectives.

5.2 Toward Rich Listening Experiences on Auditory Websites

Our findings point to promising research directions for auditory websites. First, future auditory website design should consider incorporating the design considerations emphasized by our professional sound designers, alongside the current focus on interactivity and functionality, in pursuit of richer, more enjoyable auditory website experience.

Second, the specific sound design ideas from our study could serve as helpful resources for future research to explore and experiment with—perhaps by developing higher-fidelity auditory website prototypes that instantiate the ideas identified in this paper to be evaluated by a diverse range of end users. We summarize the list of ideas below:

- *Functional and aesthetic non-speech sounds*: Sounds that are representative of the style, feeling, and properties of specific content (e.g., music, ambient sounds) may be useful for complementing spoken words for a more intuitive and navigable

interaction. The aesthetic quality of these sound elements is likely important.

- *Rich auditory icons*: Earcons and spearcons may be useful in place of spoken words for frequently occurring events to improve efficiency (e.g., transition to a new section, response to user actions) while contributing to the styling of the audio experience.
- *A sense of connection*: Audio clips that listeners are already familiar with may induce a sense of connection between users and specific website content, such as well-known commercial music and everyday sounds.
- *Pattern disruptions*: Variations in narration voices, soundtrack arrangement, ambient sounds, and music may be useful to sustain users' engagement and produce "pattern disruptions"—as articulated by our participants.
- *Auditory object groups*: Specifying a set of voices, background music, and ambient sounds for each information category may allow a listener to differentiate types of information, such as menu, system status, or subpages within a website.
- *Engaging navigation guide*: Although not the focus of our study, conversational interaction could be useful to support efficiency (e.g., direct access to menu items or other content).
- *Information dynamics*: Designers may be able to manipulate the perceived importance of different content by changing the volume, speed, and spatial quality of the narration voice (e.g., emphasize important content with a louder, stronger, closer voice).
- *Rich vs. minimalistic sound design*: Eliminating or reducing peripheral sound design elements could help to focus attention and prioritize comprehension for critical information (e.g., check-out page, news).
- *User agency and control*: Many opportunities exist to explore if and how end users want to control their audio experience configuration (e.g., volume, speed). Even being able to fully toggle audio styling on and off may be helpful for accessibility reasons. The ability to easily jump to or re-listen to any part of the website could also likely improve the experience of consuming information through audio (similar to "Bypass Blocks" for screen readers [73], but with more flexibility).

Third, and ultimately, should these audio techniques prove promising from an end user perspective, expansions will be needed for web design tools to more effectively support auditory design alongside visual design. Future design tools should be able to support three tasks involved in the design of auditory websites, as we observed in this study: (1) *website content configuration:* restructuring a website for better fit to an auditory format, such as by arranging sections in a linear fashion and converting visual information (e.g., pictures, information hierarchy, styles) into descriptions that can guide later sound design; (2) *prioritization of design considerations:* specifying the most important design requirements based on the website content (designers should also take specific websites' intended audiences' needs into consideration when soliciting design requirements); (3) *sound design:* composing or picking ambient sounds, music, and auditory icons, configuring synthesized voice qualities, and arranging different audio components based on previously solicited requirements.

Visual website layout and styling are typically configured through code (i.e., CSS) and graphical design tools (e.g., Adobe XD [2]). We envision future auditory website design tools to adopt a similar approach, potentially allowing sound designers to work in parallel with visual designers to create auditory styling analogous to CSS for the same website content. Users could thus consume the content either visually or through audio based on their needs. To achieve this goal, auditory markup languages will need to expand to support more refined manipulation of non-speech sounds and musicality to provide richer audio experience. In addition to markup, an audio-editing interface (e.g., Adobe Audition [1]) incorporated into a web-design tool would allow for more streamlined styling and a better gauge of the audio experience. To speed up the design process, design tools should provide templates for specific website content (e.g., tables, lists, hyperlinks) and different branding styles. Such tools should also explore ways to support designing auditory websites non-visually, both for accessibility reasons (drawing insights from [47, 57]) and because of the non-visual nature of the design activity.

A *computationally generated auditory website* could be particularly useful when resources are limited. Such an approach would need to address the three aforementioned main design tasks. For website content configuration, the key challenge would be to optimize the arrangement of sections and extract styling and graphical information. Existing research provides insights for automatically generating image descriptions [80], while machine learning algorithms such as GPT3 [56] may be helpful for classifying styling and sentiment. The system could further prompt users to provide their user experience requirements on different websites and in turn configure the output styling based on this prioritization. Based on this computational evaluation of sound design requirements, picking and arranging sound elements based on specified user experience criteria could potentially be treated as an optimization problem. To accommodate possible errors, we should allow designers to easily assess and configure the auto-generated interfaces. We encourage future studies to examine the effectiveness of this approach.

5.3 Limitations

While our research team attempted to focus participants' attention on creating a fundamentally auditory website, many were still heavily influenced by existing websites' visual design—as such, they ran the risk of attempting to exactly *translate* visual experiences into audio (i.e., sensory substitution [31]) without considering whether varying visual components would fit in a different modality or not. Further, some participants in our study also had limited experience with design for interactive interfaces, which may have biased their design ideas toward those in traditional media where the audiences passively consume information. To reduce these biases, future auditory websites should involve designers with more interaction design experience, and if possible, seek designers who primarily interact with technologies in audio. Moreover, during our study, the participants only had a short time to consider sound design ideas, and had limited technological support to make their mockups truly interactive. In turn, their usages of sound design techniques were sometimes motivated by convenience. Future studies should investigate the effectiveness of our explored sound design ideas on

fully functioning auditory websites. Finally, our study only examines designers' perspectives. While we use this study to explore new ways for sounds to present webpages and related considerations, a critically important next step is to understand how users feel about the presentations we have identified, including potential differences between sighted users and blind screen reader users, likely confirming the promise of some sound design ideas while showing that others are not effective.

6 CONCLUSION

In this work, we proposed and explored interactive sound design on auditory websites. Through a design activity and interviews with 14 professional sound designers, we identified five design considerations (aesthetics and emotion, user engagement, audio clarity, information dynamics, and interactivity) that could be improved through sound design on auditory websites as well as a set of specific sound design strategies to support these considerations. We presented promising research directions for future auditory websites and encourage more attention to sound design for auditory interfaces.

ACKNOWLEDGMENTS

This work was funded in part by Mozilla and Microsoft Research.

REFERENCES

[1] Adobe. 2021. *Adobe Audition. A professional audio workstation.* Retrieved September 3, 2021 from https://www.adobe.com/products/audition.html

[2] Adobe. 2021. *Adobe XD. Design like you always imagined.* Retrieved September 3, 2021 from https://www.adobe.com/products/xd.html

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

[4] R. Altman and P.R. Altman. 1992. *Sound Theory, Sound Practice.* Routledge. https://books.google.com/books?id=k_jcyAEACAAJ

[5] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (April 2019), 28 pages. https://doi.org/10.1145/3311956

[6] Apple. 2021. *GarageBand for Mac. Incredible music. In the key of easy.* Retrieved September 3, 2021 from https://www.apple.com/mac/garageband/

[7] Apple. 2021. *Logic Pro. Ridiculously powerful. Seriously creative.* Retrieved September 3, 2021 from https://www.apple.com/logic-pro/

[8] Barry Arons. 1997. SpeechSkimmer: A System for Interactively Skimming Recorded Speech. *ACM Trans. Comput.-Hum. Interact.* 4, 1 (March 1997), 3–38. https://doi.org/10.1145/244754.244758

[9] Stephen Barrass. 2012. The Aesthetic Turn in Sonification Towards a Social and Cultural Medium. *AI & society* 27 (05 2012), 177–181. https://doi.org/10.1007/s00146-011-0335-5

[10] Stephen Barrass. 2015. Leonardo Special Section: Practice-Based Research and New Interfaces for Musical Expression: An Annotated Portfolio of Research through Design in Acoustic Sonification. *Leonardo* 49 (07 2015), 498–499. https://doi.org/10.1162/LEON_a_01116

[11] Stephen Barrass and Paul Vickers. 2011. Sonification Design and Aesthetics. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 7, 145–171. http://sonification.de/handbook/chapters/chapter7/

[12] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. https://doi.org/10.1145/3264901

[13] Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. More than Meets the Eye: A Survey of Screen-Reader Browsing Strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*. Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/1805986.1806005

[14] Jennifer L. Bowie. 2012. Sound Usability? Usability Heuristics and Guidelines for User-Centered Podcasts. *Commun. Des. Q. Rev* 13, 2 (June 2012), 15–24. https://doi.org/10.1145/2424840.2424841

[15] Danielle Bragg, Katharina Reinecke, and Richard E. Ladner. 2021. Expanding a Large Inclusive Study of Human Listening Rates, Vol. 14. Association for Computing Machinery, New York, NY, USA, Article 12, 26 pages. https://doi.org/10.1145/3461700

[16] Stacy M. Branham and Antony Rishin Mukkath Roy. 2019. Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 446–458. https://doi.org/10.1145/3308561.3353797

[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[18] Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. 1993. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 222–227. https://doi.org/10.1145/169059.169179

[19] Hayden Brown. 2021. *Upwrok. The world's work marketplace.* Retrieved September 3, 2021 from https://www.upwork.com/about/

[20] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 223 (Nov. 2019), 19 pages. https://doi.org/10.1145/3359325

[21] Simon Carlile. 2011. Psychoacoustics. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 3, 41–61. http://sonification.de/handbook/chapters/chapter3/

[22] Michel Chion. 1994. *Audio-Vision: Sound on Screen* (14th. ed.). Columbia University Press, New York, NY.

[23] Dasom Choi, Daehyun Kwak, Minji Cho, and Sangsu Lee. 2020. "Nobody Speaks That Fast!" An Empirical Study of Speech Rate in Conversational Agents for People with Vision Impairments. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376569

[24] Kc Collins. 2008. *Game Sound; An Introduction to the History, Theory and Practice of Video Game Music and Sound.* https://doi.org/10.7551/mitpress/7909.001.0001

[25] Millicent Cooley. 1998. Sound + image in computer-based design: learning from sound in the arts.

[26] Eric Corbett and Astrid Weber. 2016. What Can I Say? Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 72–82. https://doi.org/10.1145/2935334.2935386

[27] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What Can i Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 43, 12 pages. https://doi.org/10.1145/3098279.3098539

[28] Alexa Developer Documentation. 2021. *Speech Synthesis Markup Language (SSML) Reference.* Retrieved September 3, 2021 from https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html

[29] Christopher Drew. 2017. Edutaining audio: an exploration of education podcast design possibilities. *Educational Media International* 54, 1 (2017), 48–62. https://doi.org/10.1080/09523987.2017.1324360 arXiv:https://doi.org/10.1080/09523987.2017.1324360

[30] Alistair D. N. Edwards. 2011. Auditory Display in Assistive Technology. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 17, 431–453. http://sonification.de/handbook/chapters/chapter17/

[31] Tayfun Esenkayaa and Michael J Proulxb. 2016. Crossmodal processing and sensory substitution: Is "seeing" with sound and touch a form of perception or cognition? *integration* 56, 3 (2016), 640–662.

[32] Sam Ferguson, William L. Martens, and Densil Cabrera. 2011. Statistical Sonification for Exploratory Data Analysis. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 8, 175–196. http://sonification.de/handbook/chapters/chapter8/

[33] Fiverr. 2021. *Find the perfect freelance services for your business Effects.* Retrieved September 3, 2021 from https://www.fiverr.com/

[34] Freesound. 2019. *Freesound.* Retrieved September 3, 2021 from https://freesound.org/

[35] Rob Gray. 2011. Looming Auditory Collision Warnings for Driving. *Human factors* 53 (02 2011), 63–74. https://doi.org/10.1177/0018720810397833

[36] João Guerreiro and Daniel Gonçalves. 2016. Scanning for Digital Content: How Blind and Sighted People Perceive Concurrent Speech. *ACM Trans. Access. Comput.* 8, 1, Article 2 (Jan. 2016), 28 pages. https://doi.org/10.1145/2822910

[37] D. Halpern, Randolph Blake, and James Hillenbrand. 1986. Psychoacoustics of a chilling sound. *Attention Perception & Psychophysics* 39 (03 1986), 77–80. https://doi.org/10.3758/BF03211488

[38] Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). 2011. *The Sonification Handbook*. Logos Publishing House, Berlin, Germany. 1–586 pages. http://sonification.de/handbook

[39] IBM. 2021. *Watson Speech to Text.* Retrieved September 3, 2021 from https://www.ibm.com/cloud/watson-speech-to-text

[40] Frankie James. 1998. Lessons from Developing Audio HTML Interface. In *Proceedings of the Third International ACM Conference on Assistive Technologies* (Marina del Rey, California, USA) *(Assets '98)*. Association for Computing Machinery, New York, NY, USA, 27–34. https://doi.org/10.1145/274497.274504

[41] Philipp Kirschthaler, Martin Porcheron, and Joel E. Fischer. 2020. What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 9, 9 pages. https://doi.org/10.1145/3405755.3406119

[42] J. Lazar, Aaron Allen, Jason Kleinman, and C. Malarkey. 2007. What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. *International Journal of Human–Computer Interaction* 22 (2007), 247 – 269.

[43] Sara Lenzi and Paolo Ciuccarelli. 2020. Intentionality and design in the data sonification of social issues. *Big Data & Society* 7 (07 2020), 205395172094460. https://doi.org/10.1177/2053951720944603

[44] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[45] Darren Lunn, Simon Harper, and Sean Bechhofer. 2011. Identifying Behavioral Strategies of Visually Impaired Users to Improve Access to Web Content. *ACM Trans. Access. Comput.* 3, 4, Article 13 (April 2011), 35 pages. https://doi.org/10.1145/1952388.1952390

[46] Yossi Matias. 2020. *Easier access to web pages: Ask Google Assistant to read it aloud.* Retrieved September 3, 2021 from https://blog.google/products/assistant/easier-access-web-pages-let-assistant-read-it-aloud/

[47] Keenan R. May, Brianna J. Tomlinson, Xiaomeng Ma, Phillip Roberts, and Bruce N. Walker. 2020. Spotlights and Soundscapes: On the Design of Mixed Reality Auditory Environments for Persons with Visual Impairment. *ACM Transactions on Accessible Computing* 13, 2 (Aug 2020), 1–47. https://doi.org/10.1145/3378576

[48] Josh H. McDermott. 2012. Chapter 10 - Auditory Preferences and Aesthetics: Music, Voices, and Everyday Sounds. In *Neuroscience of Preference and Choice*, Raymond Dolan and Tali Sharot (Eds.). Academic Press, San Diego, 227–256. https://doi.org/10.1016/B978-0-12-381431-9.00020-6

[49] Emma Murphy, Ravi Kuber, Graham McAllister, Philip Strain, and Wai Yu. 2008. An empirical investigation into the difficulties experienced by visually impaired Internet users. *Universal Access in the Information Society* 7 (04 2008), 79–91. https://doi.org/10.1007/s10209-007-0098-4

[50] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3173574.3173580

[51] Simone Natale and Henry Cooke. 2020. Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. *Media, Culture & Society* (2020). https://doi.org/10.1177/0163443720983295 arXiv:https://doi.org/10.1177/0163443720983295

[52] Michael Nees. 2019. Eight Components of a Design Theory of Sonification. 176–183. https://doi.org/10.21785/icad2019.048

[53] John G. Neuhoff. 2011. Perception, Cognition and Action in Auditory Display. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 4, 63–85. http://sonification.de/handbook/chapters/chapter4/

[54] Patrick Ng and Keith Nesbitt. 2013. Informative Sound Design in Video Games. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death* (Melbourne, Australia) *(IE '13)*. Association for Computing Machinery, New York, NY, USA, Article 9, 9 pages. https://doi.org/10.1145/2513002.2513015

[55] Donald A. Norman. 2002. *The Design of Everyday Things.* Basic Books, Inc., USA.

[56] OpenAI. 2021. *GPT-3 Powers the Next Generation of Apps.* Retrieved September 3, 2021 from https://openai.com/blog/gpt-3-apps/

[57] William Christopher Payne, Alex Yixuan Xu, Fabiha Ahmed, Lisa Ye, and Amy Hurst. 2020. How Blind and Visually Impaired Composers, Producers, and Songwriters Leverage and Adapt Music Technology. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 35, 12 pages. https://doi.org/10.1145/3373625.3417002

[58] Anne Marie Piper, Robin Brewer, and Raymundo Cornejo. 2017. Technology Learning and Use among Older Adults with Late-Life Vision Impairments. 16, 3 (Aug. 2017), 699–711. https://doi.org/10.1007/s10209-016-0500-1

[59] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 12 pages. https://doi.org/10.1145/3173574.3174214

[60] Iv Ramakrishnan, Vikas Ashok, and Syed Masum Billah. 2017. Non-visual Web Browsing: Beyond Web Accessibility, Vol. 10278. 322–334. https://doi.org/10.1007/978-3-319-58703-5_24

[61] Readaloud. 2019. *Read Aloud. Voice enabling the web.* Retrieved September 3, 2021 from https://readaloud.app/

[62] Stephen Roddy. 2020. Using Conceptual Metaphors to Represent Temporal Context in Time Series Data Sonification. *Interacting with Computers* 31, 6 (02 2020), 555–576. https://doi.org/10.1093/iwc/iwz036 arXiv:https://academic.oup.com/iwc/article-pdf/31/6/555/33525106/iwz036.pdf

[63] Stephen Roddy and Dermot Furlong. 2014. Embodied Aesthetics in Auditory Display. *Organised Sound* 19, 1 (2014), 70–77. https://doi.org/10.1017/S1355771813000423

[64] Stephen Roddy and Dermot Furlong. 2015. Sonification listening: An empirical embodied approach. In *Presented at the 21st International Conference on Auditory Display (ICAD2015), July 6-10, 2015, Graz, Styria, Austria.* (Graz, Styria, Austria). Graz, Styria, Austria. http://hdl.handle.net/1853/54125

[65] Eda Sayin, Aradhna Krishna, Caroline Ardelet, Gwenaëlle Briand Decré, and Alain Goudey. 2015. "Sound and safe": The effect of ambient sound on the perceived safety of public spaces. *International Journal of Research in Marketing* 32, 4 (2015), 343–353. https://doi.org/10.1016/j.ijresmar.2015.06.002

[66] Kristen M. Scott, Simone Ashby, and Julian Hanna. 2020. "Human, All Too Human": NOAA Weather Radio and the Emotional Impact of Synthetic Voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3313831.3376338

[67] Stefania Serafin, Karmen Franinović, Thomas Hermann, Guillaume Lemaitre, Michal Rinott, and Davide Rocchesso. 2011. Sonic Interaction Design. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 5, 87–110. http://sonification.de/handbook/chapters/chapter5/

[68] Avid Technology. 2021. *Pro Tools. Music Software for Everyone.* Retrieved September 3, 2021 from https://www.avid.com/pro-tools

[69] Lars Udsen and Anker Helms Jørgensen. 2005. The aesthetic turn: Unravelling recent aesthetic approaches to human-computer interaction. *Digital Creativity* 16 (01 2005). https://doi.org/10.1080/14626260500476564

[70] Paul Vickers. 2011. Sonification for Process Monitoring. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 18, 455–491. http://sonification.de/handbook/chapters/chapter18/

[71] Markel Vigo and Simon Harper. 2013. Coping tactics employed by visually disabled users on the web. *International Journal of Human-Computer Studies* 71, 11 (2013), 1013–1025. https://doi.org/10.1016/j.ijhcs.2013.08.002

[72] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. VERSE: Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. ACM, New York, NY, USA, 414–426. https://doi.org/10.1145/3308561.3353773

[73] W3C. 2016. *Bypass Blocks: Understanding SC 2.4.1.* Retrieved September 3, 2021 from https://www.w3.org/TR/UNDERSTANDING-WCAG20/navigation-mechanisms-skip.html

[74] Bruce N. Walker and Gregory Kramer. 2005. Mappings and Metaphors in Auditory Displays: An Experimental Assessment. *ACM Trans. Appl. Percept.* 2, 4 (Oct. 2005), 407–412. https://doi.org/10.1145/1101530.1101534

[75] Bruce N. Walker and Michael A. Nees. 2011. Theory of Sonification. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Publishing House, Berlin, Germany, Chapter 2, 9–39. http://sonification.de/handbook/chapters/chapter2/

[76] Zapsplat. 2019. *Free Sound Effects.* Retrieved September 3, 2021 from https://www.zapsplat.com/

[77] Dongsong Zhang, Lina Zhou, Judith O. Uchidiuno, and Isil Y. Kilic. 2017. Personalized Assistive Web for Improving Mobile Web Browsing and Accessibility for Visually Impaired Users. *ACM Trans. Access. Comput.* 10, 2, Article 6 (April 2017), 22 pages. https://doi.org/10.1145/3053733

[78] Lotus Zhang, Lucy Jiang, Nicole Washington, Augustina Ao Liu, Jingyao Shao, Adam Fourney, Meredith Ringel Morris, and Leah Findlater. 2021. Social Media through Voice: Synthesized Voice Qualities and Self-Presentation, Vol. 5. Association for Computing Machinery, New York, NY, USA, Article 161, 21 pages. https://doi.org/10.1145/3449235

[79] Xu Zhang, Meihui Ba, Jian Kang, and Qi Meng. 2018. Effect of soundscape dimensions on acoustic comfort in urban open public spaces. *Applied Acoustics*

133 (2018), 73–81.

[80] Xiaoyi Zhang, Lilian de Greef, Amanda Swearngin, Samuel White, Kyle Murray, Lisa Yu, Qi Shan, Jeffrey Nichols, Jason Wu, Chris Fleizach, Aaron Everitt, and Jeffrey P Bigham. 2021. Screen Recognition: Creating Accessibility Metadata for Mobile Applications from Pixels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 275, 15 pages. https://doi.org/10.1145/3411764.3445186