It's All About the Message! Visual Experience is a Precursor to Accurate Auditory Interaction

Simon Harper
School of Computer Science, University of Manchester, Manchester, UK
simon.harper@manchester.ac.uk

Sukru Eraslan
Middle East Technical University, Northern Cyprus Campus, Kalkanlı, Güzelyurt, Mersin 10 Turkey
seraslan@metu.edu.tr

Yeliz Yesilada
Middle East Technical University, Northern Cyprus Campus, Kalkanlı, Güzelyurt, Mersin 10 Turkey
yyeliz@metu.edu.tr

ABSTRACT

Anecdotal evidence suggests there is a disjoint between the interaction experiences of sighted and visually disabled web users. However, we propose the converse and suggest that this disjoint is created by the lack of understanding of the interplay between the two domains. Current research shows that there is one single locus of attention at a given time in the context of web interaction, and therefore sighted users form a serialisation of the things they look at and pay attention - an exemplar of which can be seen in eye movement sequences of users. We also suggest that web designers have a narrative in mind to be experienced by users, and they create a visual sequence they wish their audience to perceive for supporting this narrative. However, this sequence is typically lost when we move from visual presentations to auditory ones. Current audio interactions centre around page linearisation based on the sequence of the underlying source code. This linearisation typically falls short of the kind of comprehensive interaction which can be expected in the visual domain. In this paper, we use an eye tracking dataset to illustrate that the linearisation of web page component based on the underlying source code differs from what is experienced by sighted users. We then show that the web experience of visually disabled users can be improved by re-ordering the most commonly used web page components based on the order in which they are used. We also suggest that it is critical to conduct formative experimentation with sighted users to establish a visual narrative and serialisation, thereby informing the design of the auditory conversation.

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** User studies; Usability testing; Laboratory experiments; •**Information systems** → **World Wide Web;** Web interfaces;

W4A '19, May 13–15, 2019, San Francisco, CA, USA

DOI: https://doi.org/10.1145/3315002.3317554

Keywords

Eye Tracking, Scanpath Trend Analysis, STA, Experiential Transcoding, Web Pages, Visual Elements

1. INTRODUCTION

The purpose of art and design is to alter the perceptions of viewers by manipulating visual components. The main way to accomplish this purpose is to create a narrative in which the work refers to or a narrative which is created by moving through the visual components of the work in sequence. The sequence of the visual components which form the overall narrative is important or more important than the aesthetic alone. Most art is narrative and depict stories from religion, myth, legend, history and literature. Even though formalist ideas in modern art have resulted in narrative being frowned upon, the visual narrative is still present in the coded references to political or social issues, or to events in the artist's life.

The principles of the visual narrative are all captured and used in web design. Web designers use colours, text styles, font sizes, images and animations to alter the perceptions of sighted users and take them on a narrative journey through a web page [20, 9]. In this way, these users are unknowingly or unconsciously involved with the message of a website, ergo the narrative, via the serialisation of the visual components [13, 9]. This is supported by current research which shows that there is one single locus of attention at a given time in the context of web interaction, and therefore sighted users interact with web pages by using their components in a visual order [17]. This visual interaction can also be seen from eye movement sequences of users. As an example, Figure 1 shows eye movements of a particular user on the home page of the Apple website. The circles represent the fixations where the eyes become relatively immobile. The size of a circle is directly proportional to the duration of its corresponding fixation. Computational models have also been successfully employed in computer graphics to segment images into their regions which are most likely to catch the attention of users [10]. However, eye tracking studies show that the most visually obvious page features do not always catch the attention of users [11].

Anecdotal evidence suggests there is a disjoint between the interaction experiences of sighted and visually disabled web users. However, we propose that there is a direct linkage between the experiences of these users. In this paper,

Figure 1: Eye movements of a particular user on the home page of the Apple website which is segmented into its visual components by using the extended VIPS algorithm [1]

we wish to build a case that we can inform the desired outcome of the cognitive experiences of visually disabled users by understanding the cognitive experiences of sighted users, and destroy the common misconception which supports the anecdotal disjoint between sighted and visually disabled web users. We hypothesise that an understanding of the visual experience of sighted users is a precursor to any comprehensive design of auditory web-based interaction. While it is true that web accessibility relies on technical aspects of joining web content, user agents, and assistive technologies, we contend that web comprehension relies on cognition. For both user groups, the cognitive message is the same, but the perception of that message is different.

Current visual to auditory mapping frameworks mainly attempt to linearise web pages based on the underlying code structure with certain and not so intelligent attempts at understanding the semantic nature of the block structure of the Hypertext Mark-up Language (HTML or XHTML), and take the form of top-left to bottom-right serialisation. Hereafter, we refer to this as linear order. There have been no successful attempts at automatic linearisation based on the combination of visual rendering and a knowledge of visual behaviour of humans. However, this combination is critical in considering a web page as a document where visual rendering, user behaviour and interaction supported by this rendering is considered together. Hereafter, we refer to this as visual order.

In the rest of the paper, we firstly explain conventional linearisation and transcoding techniques (Section 2). We then use a dataset from an existing eye tracking study to show that there is no correlation between the linear order and visual order of web page components, and illustrate that when web pages are transcoded by re-ordering their components based on the visual order, visually disabled users would access these components with significantly less time without being distracted by other components (Section 3). After that, we discuss the problems which can be raised when the visual components of web pages are serialised based on the order in which they are used (Section 4). We also discuss this serialisation with regard to generalisable research methods (Section 4.1) and finally conclude with a round-up of our rationale (Section 5).

Contributions: Even though currently available screen readers could avoid syntactical linearisation by allowing direct access to some specific types of web page elements (e.g., links), they still rely on the linear order, in particular the linear order of these elements, in the source code. Therefore, screen readers consider the source code order and do not take the visual order into consideration. To the best of our knowledge, the work presented in this paper is the first work for investigating transcoding of web pages for visually disabled users based the visual order constructed from eye movement sequences of real web pages.

2. TRANSCODING

If we wish to move beyond simple technical accessibility, then we must work towards 'Interface Equivalence' which means an experience that is cognitively and temporally similar across all users with regard to both interface and information: perception, navigation, orientation, and comprehension. By understanding the serialisation of the visual sequence, we can understand the serialisation of the auditory sequence to re-produce an equivalent user experience. Currently, this is not the case due to the way that web pages are presented for visually disabled users. Web pages for these users are currently transformed into a linear sequence based on the underlying structure of their source code.

Web pages can also be transcoded with different techniques by adapting their content so that they can be viewed on any of the increasingly diverse devices found on today's market. Transcoding has been used for a number of years in the context of making incomplete or badly written hypertext accessible to visually disabled users and their accessibility technologies. Some of the transcoding techniques are as follows: adding a skip link, ranking and reordering components, removing irrelevant components, etc. Further information about these techniques can be found in [3].

While transcoding enables us to provide better technical information to user agents, it does not, on the whole, enable us to provide an equivalent end-user experience because the user agents do not 'understand' the order in which the visual narrative should be serialised. We can see this serialisation in more detail when we examine trending paths in eye movements of users. Therefore, we suggest to apply transcoding techniques by taking trending paths into consideration for making web pages more accessible and usable for visually disabled users. The following section provides an empirical work to support our suggestion.

3. EYE TRACKING APPROACH

Eye tracking technologies are now increasingly used in studies that analyse user behaviours in searching on the web or to reveal possible usability and accessibility problems [15, 8, 12, 4]. While we are reading, looking at a scene or searching for an component, our eyes do not generally

move smoothly over the visual field but they make continuous movements called saccades. Between these saccades, our eyes become relatively stable at certain points called fixations and the sequence of these fixations shows our scanpath (see Figure 1). Fixations are the periods that indicate where our eyes pay more attention, hence which component is viewed. Therefore, tracking eye movements has now become a valuable way of understanding how people allocate their visual attention. As it is suggested in our previous work [19], eye tracking data has the potential to be used as the guide for understanding the visual ordering of a web page – mainly the ordering used by sighted users to interact with the page.

There have been a number of studies examining eye movements during both browsing and searching on the web. We have re-analysed the data from one of these studies to understand and support our research argument.

3.1 Dataset

The dataset used in this analysis is from our previous eye tracking study. The detailed description of the study can be found in [7] and its brief summary is provided below.

Participants: Twenty male and twenty female were involved in the study which was conducted at the University of Manchester and Middle East Technical University Northern Cyprus Campus. The majority of the participants were students along with some academic and administrative staff.

Equipment and Materials: The eye movements of the participants were recorded with Tobii T60 eye tracker integrated into a 17" monitor. The home pages of the following popular websites were used which had varying degree of visual complexity determined with the ViCRAM tool [16]: Apple (Low Complexity), Babylon (Low Complexity), AVG (Medium Complexity), Yahoo (Medium Complexity), Godaddy (High Complexity) and BBC (High Complexity).

Procedure: The participants were asked to perform some searching and browsing tasks on the web pages in a random order. There was no specific objective for the browsing tasks where the participants spontaneously viewed the pages without clicking any links. In contrast, they required to find specific information or items on the pages for the searching tasks. For example, on the home page of the Apple website, they were asked to locate the link that allows to watch the TV ads relating to iPad mini and a link labelled iPad on the main menu. These tasks were developed based on the G. Marchionini's search activities model [14] consisting three task categories: look up, learn and investigate. This model is one of the most popular models on the categorisation of tasks on the web. The browsing task is relevant to serendipitous browsing which is part of the investigate group whereas the searching task is relevant to fact finding which is part of the look up group.

3.2 Research Questions

To support our research argument, we focused on the following two research questions:

1. *Is there any correlation between the linear order and visual order of web page components?*

   This question aims to investigate whether the linear order is correlated with the visual order (i.e., the order of the components in which they are used). In other words, this question aims to investigate whether the linearisation of web page components based on the source code makes the experience of visually-disable users equivalent to the experience of sighted users. Our previous work [19, 2] suggests that the linear order and the visual order do not correlate, however there has been no data-driven work to support this observation.

2. *Do visually disabled users access the most commonly used components on web pages with significantly less time when these components are re-ordered based on their visual orders?*

   This question aims to investigate whether the web experience of visually disabled users is improved by re-ordering the most commonly used components on web pages based on the order in which they are used. We expected that visually disabled users would need significantly less time to access and use the most commonly used components on web pages after transcoding.

Our dataset allowed us to investigate our research questions by using different real web pages with varying level of visual complexities and real web users.

3.3 Methodology

In order to investigate these research questions, we firstly needed to identify visual and linear orders of the web page components. We started with an approach to detect the components of the web pages and we used an existing tool to identify the linear order of these components. For the visual order, we used the data collected in the eye tracking study explained above and we generated a trending path of the participants of that study on each page. The trending paths were represented in terms of the web page components. Therefore, we were able to compare the visual and linear orders as they were both represented in terms of the web page components. Regarding the second research question, we mainly compared the time to access the components in the visual and linear orders. Below we explain each step in full detail.

The extended Vision-based Page Segmentation (VIPS) algorithm was used for identifying visual components of the web pages as it automatically identifies visual components and correlates them with the underlying source code of web pages such that these components can then be used for further processing [1]. The VIPS algorithm generates a tree of components where there are more and smaller components in the deeper levels. The fifth level was used in this analysis as a user study conducted by [1] suggested this level as the most preferred level by users.

The Scanpath Trend Analysis (STA) algorithm was used to identify the trending components and the order of these components on each page by analysing the eye movements of the users for the browsing and searching tasks. The STA algorithm identifies a trending path for multiple users on a particular web page in terms of its visual components. It takes a series of fixations for each user on a web page and the visual components of that page, and then finds the corresponding component for each fixation to create the individual scanpaths in terms of the components. After that, the algorithm analyses the individual scanpaths to discover

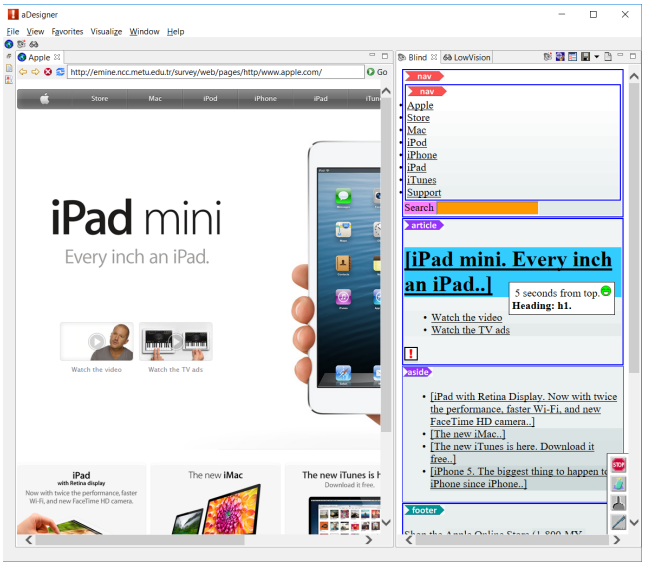

Figure 2: The home page of the Apple website on the aDesigner tool

the components to be in the trending path by selecting the components shared by all users and the components that get at least the same attention as the shared components in terms of the total fixation durations and total fixation counts. Finally, it puts the selected components into the trending path based on their overall positions in the individual scanpaths. Trending components can be seen more than once in a trending path. In this case, the repetitions were excluded from the trending path, and the order of each component was noted as a visual order of that component.

The aDesigner[1] tool was then used to linearise the web page components, identify the linear order of the trending components and determine access times to each visual component in seconds. As an example, Figure 2 shows how this tool works with the home page of the Apple website. This tool is a well-formed and well-accepted disability simulator which helps designers ensure that their websites are accessible and usable by visually disabled users. It is developed based on the experiments conducted with real visually-disabled users to reflect the experiences of real users. As this simulator is also open-sourced, its results are replicable [18]. Since there were some problems with the linearisation of the home page of the Godaddy website due to its source code, the page was excluded from the further analysis.

Once the linear order and visual order of the trending components were identified, a correlation analysis was then conducted for each page for the browsing and searching tasks to investigate whether there is a correlation between the linear order and visual order of these components.

The web pages were also transcoded in two different ways. In both of these ways, the trending components were re-ordered based on their visual orders and the other components were removed. In the first version, the components were only re-ordered, but in the second version the components were re-ordered and also a specific heading was assigned to each element. The access times to the trending components in the transcoded versions were also determined by using the aDesigner tool.

A paired dependent T-test or its non-parametric alternative Wilcoxon signed rank test (when the data was not normally distributed) was used to investigate whether the transcoded versions of the web pages allow accessing the trending components with significantly less time for the browsing and searching tasks.

3.4 Results

Table 1 and Table 2 illustrate the visual order and linear order of the trending components of the five web pages along with their access times in the original version of the web pages and their two transcoded versions in seconds for the browsing and searching tasks respectively. As an example, Figure 3 visualises the linear order and visual order of the components of the home page of the Apple website where the white circles, black circles and grey circles show the linear order, the visual order for the browsing task and the visual order for the searching task respectively.

Figure 3: The linear order and visual order of the components of the home page of the Apple website where the white circles, black circles and grey circles show the linear order, the visual order for the browsing task and the visual order for the searching task respectively

Due to the limited number of the trending components, a correlation analysis could not be conducted on the home page of the AVG website. However, as we expected, there was no correlation between the linear order and visual order of the trending elements on the rest of the pages for both the browsing and searching tasks, apart from the home page of the Babylon website for the searching task. The detailed results of our correlation analyses are shown in Table 3.

Table 4 shows the descriptive statistics of access times to the trending components in the original version of the web pages and their two transcoded versions in seconds.

Our statistical analyses show that the first transcoded ver-

[1]http://www.eclipse.org/actf/downloads/tools/aDesigner/

Table 1: Access times in seconds for the visual and linear orders of the web page components for the browsing task in the original (Orig.) and the two transcoded versions in seconds (Ver. 1: components ordered in visual order, Ver. 2: components ordered in visual order and has headings)

Page	Area	Visual Order	Linear Order	Orig.	Ver. 1	Ver. 2
Apple	C	1	3	5	0	0
	F	2	6	11	2	3
	I	3	9	14	2	4
	H	4	8	13	6	9
	B	5	1	0	4	11
	G	6	7	7	12	12
	E	7	5	9	18	17
Babylon	M	1	13	16	0	0
	H	2	8	5	0	1
	L	3	12	13	4	6
	I	4	9	6	6	9
	R	5	18	25	8	16
	Q	6	17	17	20	19
	S	7	19	36	28	21
	P	8	16	10	34	23
AVG	G	1	7	15	0	0
	I	2	9	32	14	9
Yahoo	I	1	9	5	0	0
	J	2	10	30	18	20
	G	3	7	31	53	29
BBC	L	1	12	11	0	0
	S	2	19	83	194	7
	T	3	20	101	213	9
	P	4	14	14	5	11

sion allows to access the trending components with significantly less time for the searching tasks with medium effect size (Dependent T-Test, t = 3.089, p = 0.003, df = 25, d = 0.61, one-tailed), but this is not the case for the browsing tasks (Wilcoxon, Z = -1.035, p = 0.150, r = 0.15, one-tailed). However, the second transcoded version allows to access the trending components with significantly less time for both the browsing (Wilcoxon, Z = -2.601, p = 0.005, r = 0.38, one-tailed) and searching (Wilcoxon, Z = -2.436, p = 0.008, r = 0.34, one-tailed) tasks with medium effect size. Based on these results, we can suggest that when web pages are transcoded by considering the visual order of web page components, they can become more accessible for visually disabled users as they can access trending elements more quickly.

3.5 Discussion

The first research question aimed to investigate whether or not the linear order is correlated with the visual order. The correlation was only found on the home page of the Babylon website for the searching tasks, but it was obviously caused by the given searching task. The task on the page was to locate the link for downloading the free version of Babylon and read the names of other products. To complete this task, the participants needed to locate the elements in their linear orders. Apart from this case, the results support our expectation.

The second question aimed to investigate whether or not visually disabled users can access trending components on web pages with significantly less time after transcoding web pages based on the visual order of trending components. The overall results support our expectation, but there was no

Table 2: Access times in seconds for the visual and linear orders of the web page components for the searching task in the original (Orig.) and the two transcoded versions in seconds (Ver. 1: components ordered in visual order, Ver. 2: components ordered in visual order and has headings)

Page	Areas	Visual Order	Linear Order	Orig.	Ver. 1	Ver. 2
Apple	I	1	9	14	0	0
	C	2	3	5	3	4
	F	3	6	11	6	8
	H	4	8	13	6	9
	E	5	5	9	7	11
	B	6	1	0	4	14
	G	7	7	7	12	12
Babylon	M	1	13	16	0	0
	H	2	8	5	0	1
	I	3	9	6	4	6
	R	4	18	25	8	12
	P	5	16	10	20	17
	N	6	14	9	9	19
	Q	7	17	17	10	21
	S	8	19	36	18	25
AVG	G	1	7	15	0	0
	I	2	9	32	14	9
Yahoo	J	1	10	30	0	0
	I	2	9	5	4	13
	G	3	7	31	24	28
BBC	L	1	12	11	0	0
	P	2	14	14	5	7
	R	3	18	26	50	11
	S	4	19	83	77	13
	T	5	20	101	95	15
	N	6	16	18	7	17

Table 3: The correlation analyses of the visual order and linear order of the trending components of the five web pages for the browsing and searching tasks [r: Correlation coefficient, n: sample size, *: statistical significance]

Task	Page	Test	r	n	p
Browsing	Apple	Pearson	-0.301	7	0.512
	Babylon	Pearson	0.718	8	0.045*
	AVG	NA	NA	NA	NA
	Yahoo	Pearson	-0.982	3	0.121
	BBC	Pearson	0.676	6	0.140
Searching	Apple	Pearson	0.000	7	1.000
	Babylon	Pearson	0.704	8	0.051
	AVG	NA	NA	NA	NA
	Yahoo	Pearson	-0.655	3	0.546
	BBC	Pearson	0.234	4	0.766

significant difference between the access times in the original version and the first transcoded version for the browsing task. It was mainly caused by the home page of the BBC website as the headings were not appropriately used on the page and therefore a large amount of time was needed to access some trending components (see Table 1). When we provide a specific heading for each trending component, we achieved a statistically significant difference as mentioned above.

Web page transcoding based on understanding user experiences is referred to as "Experiential Transcoding" in the literature [19]. The results presented here show that this kind

Table 4: The descriptive statistics of access times to the trending components in the original version of the web pages and their two transcoded versions in seconds [M: Mean, MD: Median, SD: Standard Deviation, T. Ver. 1: Transcoded Version 1, T. Ver. 2: Transcoded Version 2]

Task	Version	M	MD	SD	Min	Max
Browsing	Original	21.21	13.50	23.94	0	101
	T. Ver. 1	26.71	6.00	56.04	0	213
	T. Ver. 2	9.83	9.00	8.39	0	29
Searching	Original	21.12	14.00	23.06	0	101
	T. Ver. 1	14.73	6.50	23.62	0	95
	T. Ver. 2	10.46	11.00	7.90	0	28

of transcoding would improve the web experience of visually disabled users by allowing them to access the most commonly used components directly without being distracted by other inappropriate components. In the literature, previously there has been an attempt to transcode web pages based on common paths of users [2]. In this paper, a different algorithm was used for detecting trending paths of participants in an eye tracking study. Although the algorithm used to identify a common path was not as successful as the STA algorithm, the results were promising [6, 5, 7]. In that study, the authors focused on task completion, system usability, information quality and interface quality. As the STA algorithm is able to provide the most representative path for multiple users, we expect to have better experiential transcoding with the STA algorithm such that it will significantly improve the web experience of visually disabled users [7]. However, further studies can be conducted to confirm this. Different algorithms can be used to identify trending paths and then these paths can be used to guide the transcoding. These alternative transcoded versions can then be compared with the aDesigner tool. Besides this, we focused on access times to web page components to measure the efficiency of experiential transcoding with the STA algorithm. Further studies can be conducted to investigate how visually disabled users are satisfied with this kind of transcoding.

In this work, we use eye tracking to guide the transcoding or for mainly understanding the visual rendering of web pages. However, further studies can be conducted to understand user behaviour in different ways. For example, some unobtrusive recordings of user behaviours can be used to detect and understand user behaviour. In particular, machine learning can be used to drive the understanding of the visual rendering of web pages.

4. FUTURE DIRECTIONS

In the visual world, we expect or anticipate that there will be parallel and competing visual components that must be in some way mapped into an auditory sequence. The order in which we experience facets of the visual work, its relationship to other aspects of the work, are both important for accurate comprehension. This comprehension is missing from current auditory presentations because they do not take the visual sequence of the presentation into account but rather conform to simplistic notions of linearisation.

We can see that visual components are serialisable at a component by component level. We also know that the locus of attention works at a component by component level and this focus is common among sighted and visually disabled users because it is based on comprehension and cognition, not perception. Understanding these key points allows us to link these two domains, and in this way drive accurate mappings to support interface equivalence and create an equivalent interactive experience.

As we have already seen in Section 3, the visual components of web pages can be serialised based on the order which they are visited to improve the web experience of visually disabled users. However, the problems can still quite legitimately occur when mapping or changing visual content:

Understanding the Participant: As it is the case with most Human-Computer Interaction (HCI) studies, the results of participants are prone to vary. If there is high variation between participants, then a trending path may not be created. However, this raises an interesting question regarding the visual design. If the variation is very high, then the visual design itself may not be appropriate for conveying the intended narrative.

Understanding the Purpose: Understanding the purpose of users is key to understanding the narrative, ergo the mapping. Browsing, task completion, and the combination of the two (this last is more common) is driven by movement in the information space and control of what to read or examine. While chance or synchronicity may have some part to play in these behaviours, users are still in control of filtering the information presented. However, since multiple factors affect this movement, an understanding of visual to auditory mapping may be obscured by explicit browsing/task completion or by implicit switching between the two.

Understanding the Design: Good design, as in good art, is not universal. There are good and bad designs, and websites which exhibit no explicit design at all. By understanding that bad, misguided, or incompetent design will direct sighted users to wrong resources by transmitting them a wrong narrative, we must also acknowledge that following this visual design in the auditory mapping will likewise give an unintended experience. While this is a problem for good auditory mapping, a strict following of the visual experience will produce an equivalent interactive auditory experience, even if that experience happens to be universally bad.

Understanding the Dynamics: Change at the component level is a key problem for auditory mapping. Static visual resources allow a fixed unchanging narrative to be created and mapped. However, this is not the case for changing visual components. Changes in context and multiple dynamic updates all compete for the attention of users and produce an incoherent cacophony if the delivery is mapped to an auditory presentation. In this case, naïve static mappings can no longer support an equivalent interactive experience. Of course, it may be that visually changing components are available for use but are shown to be either not perceived or not used by sighted participants. However, in the worst case scenario, the question of how to deal with changing content arises. While we have no fixed solutions, we consider it to be helpful to think of this updating content as a conversation between multiple

parties which needs to be orchestrated to allow accurate comprehension. However, we must also remember that if dynamic aspects are removed, then the design intention may be lost.

Understanding the Implementation: Even though experiential transcoding is promising to improve the web experience of visually disabled users, researchers should study on how it can be implemented. Web pages can be used in different ways for various reasons, so it can be difficult to decide which visual order should be taken into consideration for transcoding web pages for visually disabled users. The main tasks on the web pages can be identified by conducting a longitudinal study with web users, and the visual orders of these tasks can be then identified. After that, the visual orders can be specified within the source code of web pages to be used by assistive technology to transcode web pages for visually disabled users based on their tasks.

Understanding the Limitations: Finally, we need to understand the current limitations of our knowledge and abilities to create accurate mappings. We can make every effort to mitigate general failures or inaccuracies in the formative experiments and the mappings created from them by using triple blind techniques, removing memory effects, understanding the learning process, and distancing ourselves from a desire for a certain outcome. However, certain aspects cannot easily be overcome. Here, we mean aspects of user prediction, understanding the way a user will react to visual cues regardless of task, context, or comprehension. In these cases, we must realise that universal solutions are not yet possible.

To sum up, while visual to auditory mapping based on a model of the serialisation of visual components is eminently feasible, certain aspects must be taken into account. The model must be specific for the content under investigation (as we cannot predict user interaction with visual cues) and the user dependant aspects must be considered. It is initially best to create a static interaction model and then extend it with dynamic features, but there exists no good model of how to deal with these competing visual components. However, testing if these components are seen and used in the visual domain may suggest possible simplifications to the mapping.

4.1 Importance for Research Methods

The first common step in most web accessibility experimentation for visually disabled users is to recruit participants for formative empirical studies. The approach of starting with visually disabled users 'seems' logical, and has the misguided advantage of conforming to an advocacy of the prevailing political or social circumstances[2] by including visually disabled users from the outset. However, our argument runs deeply contrary to this view. While we support the fundamental involvement of visually disabled users, we suggest that the tendency for the view that formative studies with visually disabled participants and visual resources should be pre-eminent is misguided.

[2]Political Correctness, if you will. Here used pejoratively; as opposed to the it's true meaning, of the rejection of language, behaviour, etc., considered discriminatory or offensive, which we would subscribe to.

We assert that the first thing for any study of auditory interaction should be a study of visual interaction to form a sequence of the visual components experienced, in what order, and for how long. An understanding of how sighted users interact with multiple visual components embedded into web documents (non-linear and hyper-linked) is key because, as we have shown, the web is based on a visual interaction model. By understanding this model in the context of sighted users, we enable researchers to understand the interaction requirements needed to support users with visual disability.

This understanding is important for conventional page linearisation and transcoding techniques. It means that we need to change these methods and adopt more appropriate approaches when investigating cross-modal interaction. The first piece of the jigsaw is understanding visual interaction using eye tracking technologies. If a web page is correctly designed from a visual perspective based on eye tracking data for user trials, then each visual component can be accurately sequenced. In this way, we can establish the presentational components of the web page but not the interaction (in this case: dynamic changes + user input). We can also see what parts of the visual presentation, including dynamic components, are available for use, perceived by users, and then actually used. In this way, we are only left with the interactive input parts and the dynamic visual 'conversation' needed to be created out of the updating content that users interact with. As we can imagine, these aspects are still difficult areas to address accurately. This paper highlights a need for an experimentation to understand how an auditory presentation should be accomplished, and also supports researchers in focusing on more challenging problems, without research repetition.

Of course, there are pragmatic reasons for following this approach. Most researchers working in web accessibility, or assistive technology for visually disabled users in general, understand that soliciting participants for experimentation is very difficult. Indeed, recruiting visually disabled participants is one of the most significant hurdles to overcome. Visually disabled participants are often jealously guarded as researchers understand and value their importance and willingness to participate. There are often competing demands placed on visually disabled users when it comes to research and development work, and often this work is of interest only in the context of a better understanding of human behaviour as opposed to any clear practical contributions. Therefore, visually disabled users do not agree to the many requests for participation they receive and target their responses to more practical experimentation. In this case, we need to understand that any reduction in the requirement for visually disabled participants will be advantageous. By removing the need for initially start with formative studies of visually disabled users, we can reduce the load on participants and target experimentation in novel areas where visually disabled participants are most required.

5. CONCLUSION

We can see that the concept of parallel and serial processing with regard to vision and hearing respectively is a misconception. In reality, the serialisation occurs at the visual component level in both domains. An understanding of this point is key for good visual to auditory mappings and must occur as the first and formative study of visual

resources. Returning to our initial assertion, it is critical, for resources which have been designed visually, to conduct formative experimentation with sighted users to establish a visual narrative and serialisation, thereby informing the design of the auditory conversation.

Only by a knowledge of the narrative implicitly captured within a visual presentation, we can understand which visually salient features sighted users are drawn to, and only by this knowledge we can create accurate mappings. There are still problems to be overcome with these multi-modal mappings, specifically the generalisability of the visual serialisation of users and the unpredictable aspects of dynamic visual updates. However, the larger part of the research challenge, the serial mediation, is a mirage in reality.

In this paper, we have formed the argument that understanding the visual experience is a precursor, and is primarily important for understanding the auditory experience. This formative understanding of sighted user interaction and visual attention directly informs the design of cognitively accurate audio presentation for visually disabled user interaction. We have seen that sighted users move through visual components in particular sequence but dynamically. However, visually disabled users do not have this ability of movement and memory unless provided by technology.

By investigating the visual experiences of sighted users while interacting with visual resources which are designed to be experienced visually, we can truly understand the methodology which should be used to convert those resources into an auditory presentation which will not only be accessible but also supports an equivalent interactive experience.

6. REFERENCES

[1] M. E. Akpınar and Y. Yeşilada. *Vision Based Page Segmentation Algorithm: Extended and Perceived Success*, pages 238–252. Springer, Cham, 2013.

[2] M. E. Akpınar and Y. Yeşilada. "old habits die hard!": Eyetracking based experiential transcoding: A study with mobile users. In *Proceedings of the 12th Web for All Conference*, W4A '15, pages 12:1–12:5, New York, NY, USA, 2015. ACM.

[3] C. Asakawa and H. Takagi. Transcoding. In S. Harper and Y. Yesilada, editors, *Web Accessibility: A Foundation for Research*, Human-Computer Interaction Series, chapter 14, pages 231–260. Springer, London, 1st edition, September 2008.

[4] C. Ehmke and S. Wilson. Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1*, BCS-HCI '07, pages 119–128. British Computer Society, 2007.

[5] S. Eraslan and Y. Yesilada. Patterns in Eyetracking Scanpaths and the Affecting Factors. *Journal of Web Engineering*, 14(4&5):363–385, 2015.

[6] S. Eraslan, Y. Yesilada, and S. Harper. Identifying Patterns in Eyetracking Scanpaths in Terms of Visual Elements of Web Pages. In S. Casteleyn, G. Rossi, and M. Winckler, editors, *Web Engineering*, volume 8541 of *LNCS*, pages 163–180. Springer, 2014.

[7] S. Eraslan, Y. Yesilada, and S. Harper. Scanpath trend analysis on web pages: Clustering eye tracking scanpaths. *ACM Trans. Web*, 10(4):20:1–20:35, Nov. 2016.

[8] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky. Eye tracking in web search tasks: Design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 51–58, New York, NY, USA, 2002. ACM.

[9] R. Hoffmann and K. Krauss. A critical evaluation of literature on visual aesthetics for the web. In *Proceedings of the 2004 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, SAICSIT '04, pages 205–209, Republic of South Africa, 2004. South African Institute for Computer Scientists and Information Technologists.

[10] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998.

[11] C. Jay, R. Stevens, M. Glencross, A. Chalmers, and C. Yang. How people use presentation to search for a link: expanding the understanding of accessibility on the web. *Universal Access in the Information Society*, 6(3):307–320, Nov 2007.

[12] S. Josephson and M. E. Holmes. Visual attention to repeated internet images: Testing the scanpath theory on the world wide web. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 43–49. ACM, 2002.

[13] K. Krauss. Visual aesthetics and its effect on communication intent: a theoretical study and website evaluation. *Alternation*, 12(1a):305–329, 2005.

[14] G. Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, Apr. 2006.

[15] J. D. McCarthy, M. A. Sasse, and J. Riegelsberger. Could i have the menu please? an eye tracking study of design conventions. In E. O'Neill, P. Palanque, and P. Johnson, editors, *People and Computers XVII — Designing for Society*, pages 401–414, London, 2004. Springer London.

[16] E. Michailidou, S. Harper, and S. Bechhofer. Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, SIGDOC '08, pages 215–224, New York, NY, USA, 2008. ACM.

[17] J. Raskin. The humane interface (book excerpt). *Ubiquity*, 2000(May), May 2000.

[18] H. Takagi, C. Asakawa, K. Fukuda, and J. Maeda. Accessibility designer: Visualizing usability for the blind. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '04, pages 177–184, New York, NY, USA, 2004. ACM.

[19] Y. Yesilada, S. Harper, and S. Eraslan. Experiential transcoding: An eyetracking approach. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, W4A '13, pages 30:1–30:4, New York, NY, USA, 2013. ACM.

[20] H. Zettl. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth series in broadcast and production. Cengage Learning, 2013.