Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Sukru Eraslan
Middle East Technical University,
Northern Cyprus Campus,
99738 Kalkanlı, Güzelyurt,
Mersin 10, Turkey

Yeliz Yesilada
Middle East Technical University,
Northern Cyprus Campus,
99738 Kalkanlı, Güzelyurt,
Mersin 10, Turkey

Simon Harper
School of Computer Science,
University of Manchester,
Manchester, M13 9 PL
United Kingdom

User studies are typically difficult, recruiting enough users is often problematic and each experiment takes a considerable amount of time to be completed. In these studies, eye tracking is increasingly used which often increases time, therefore, the lower the number of users required for these studies the better for making these kinds of studies more practical in terms of economics and time expended. The possibility of achieving almost the same results with fewer users has already been raised. Specifically, the possibility of achieving 75% similarity to the results of 65 users with 27 users for searching tasks and 34 users for browsing tasks has been observed in scanpath trend analysis which discovers the most commonly followed path on a particular web page in terms of its visual elements or areas of interest (AOIs). Different approaches are available to segment or divide web pages into their visual elements or AOIs. In this paper, we investigate whether the possibility raised by the previous work is restricted to a particular page segmentation approach by replicating the experiments with two other segmentation approaches. The results are consistent with ~5% difference for the searching tasks and ~10% difference for the browsing tasks.

Open Data

The dataset used for the experiments is provided in our external online repository at http://iam-data.cs.manchester.ac.uk/data_files/33. It consists of the saved versions of the web pages used, information sheet, consent form, questionnaire and individual scanpaths in terms of the AOIs of the web pages. The repository also includes all the raw data from the experiments. Besides, the Python implementation of the Scanpath Trend Analysis (STA) algorithm can be accessed from https://github.com/SukruEraslan/sta.

Introduction

User studies play an important role in improving user experience on the web. In these studies, eye tracking has been widely used for assessing the quality of user experience on web pages (Ehmke & Wilson, 2007). It has also been used for investigating user interactions with web pages to provide some directions to improve user experience (Akpınar & Yeşilada, 2015; Brown, Jay, & Harper, 2010; Yesilada, Harper, & Eraslan, 2013). However, as

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

eye tracking researchers, we are commonly having difficulties finding users for our studies (Jay, Lunn, & Michailidou, 2008). In an eye tracking study, a researcher is required to allocate a separate session for each user, and therefore the study cannot be carried out in parallel with only one eye tracker and/or one researcher. Hence, these studies can take a significant amount of time to be completed. To deal with this issue, it is important to estimate the ideal number of users to understand and model user behaviours on the web. Even though eye tracking is increasingly used in usability studies to evaluate and improve usability of web pages (Ehmke & Wilson, 2007), existing research does not focus on the effects of the number of users on data analysis in this field. Specifically, the number of users required for analysing eye movement sequences (i.e., scanpaths), which is typically conducted based on visual elements or areas of interest (AOIs) of web pages, has not been studied in depth in the literature (Eraslan, Yesilada, & Harper, 2016a). Different approaches can be used to divide or segment web pages into their visual elements or AOIs manually or automatically (Yesilada, 2011). Researchers can apply an automated approach to discover AOIs of web pages which typically uses the source code with some heuristics. Researchers can also manually define their AOIs based on their goals. Specifically, if they are interested in certain elements on web pages (such as, advertisements), they can directly define them as their AOIs. However, there is no investigation of possible effects of AOIs on the required number of users for scanpath analysis. In the rest of the paper, we will refer to the process of identifying AOIs or visual elements on web pages as segmentation, thus the AOIs or elements will be referred to as segments.

It has been shown that there is a possibility of achieving almost the same results with fewer users, especially achieving 75% similarity to the results of 65 users with 27 users for searching tasks and 34 users for browsing tasks in scanpath trend analysis (STA) (Eraslan et al., 2016a). The STA algorithm is designed to identify the most commonly followed path as trending path on a particular web page in terms of the AOIs of the page (Eraslan, Yesilada, & Harper, 2016b). The trending path provided by the STA algorithm can be used to improve the web experience of users in constrained environments, especially visually disabled users (Eraslan et al., 2016b). In particular, web pages can be adapted for visually disabled users by making commonly used areas more accessible in the desired order such that the users can directly access

these areas with their screen readers without spending unnecessary time on clutter (Yesilada et al., 2013).

To investigate the possibility of achieving almost the same results with fewer users in scanpath trend analysis, only the extended and improved version of the Vision-based Page Segmentation (VIPS) approach was used by Eraslan et al. (2016a) to automatically segment the web pages into their AOIs (Akpınar & Yeşilada, 2013). However, the selection of a page segmentation approach may affect the possibility of achieving the same results with fewer users in scanpath analysis. Therefore, in this paper, we replicate the experiments from Eraslan et al. (2016a) with the same eye tracking dataset by using different segmentation approaches to investigate whether this possibility is restricted to a particular segmentation approach.

As an example, Figure 1 shows the Babylon page with its AOIs automatically discovered by the extended and improved version of the VIPS approach (Akpınar & Yeşilada, 2013). The VIPS approach segments web pages based on both their source code and visual representation, and relates their elements to the underlying source code (Akpınar & Yeşilada, 2013). Figure 1 also illustrates a scanpath of a particular user on the Babylon page where the circles represent the points fixated by the user and the largest circle shows the longest fixation (Tobii Technology AB, 2010). As illustrated in the figure, scanpath analysis allows the investigation of which elements catch attention and which paths are followed in terms of the elements (Eraslan et al., 2016b).
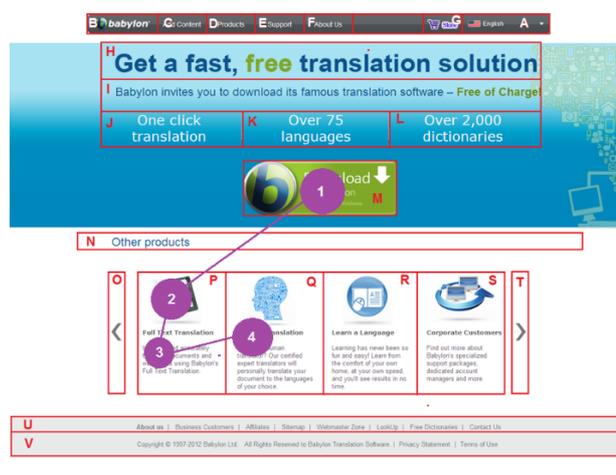

Figure 1. A user scanpath on the Babylon web page that is segmented into its AOIs by the VIPS approach

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

To investigate the validity of the possibility of achieving almost the same results with fewer users with other page segmentation approaches, we firstly reviewed the literature to find appropriate segmentation approaches for our study. However, we could find only a few public segmentation approaches. This is a common problem in computer science as highlighted by Collberg and Proebsting (2016). Thus, we could replicate the experiments with two other page segmentation approaches. As an alternative to the VIPS approach, we used the Block-o-Matic (BOM) approach for the automatic segmentation (Sanoja & Gançarski, 2014; Yesilada, 2011). The BOM approach also segments web pages based on both their source code and visual representation, but uses an alternative hybrid algorithm to do that. As researchers can also prefer to segment web pages manually based on their goals (Owens, Chaparro, & Palmer, 2011), in addition to the BOM automatic segmentation approach (Sanoja & Gançarski, 2014), we decided to apply a user driven approach to manually segment the web pages into their AOIs. For the manual segmentation, we again used the same dataset (Eraslan et al., 2016a) because the participants were asked to draw what they remember about the layout of the web pages on a piece of paper. This method is called recall. As it is explained by Johnson (2010), visual cues are very important to support users to recognise where they are. This recall approach helps us to understand which visual elements/AOIs of the pages really supported that. Those recalled by the participants were more likely to support their tasks. After the participants were asked to draw what they recall, we combined their drawings to manually segment the pages into their elements. We could also segment the web pages manually by ourselves but we used the recall approach to be more objective. If we drew the segments by ourselves, this would reflect our particular view of the segments on the pages which would not be objective. By using the recall method, we discovered how the participants generally segmented the pages in their minds and how the design of the pages supported that. Since they drew the segments based on what they remember about the pages, the drawn segments would have taken their attention, thus could be valuable in scanpath trend analysis.

The rest of this paper firstly discusses the related work, secondly explains our methodology including the STA algorithm, the eye tracking study, the segmentation approaches used and our analysis procedure, then presents the results along with their discussion and finally provides the conclusions.

Related Work

Usability studies are conducted to evaluate a particular product, such as a web page. Usability experts can observe how users interact with a particular product to complete certain tasks and/or they can examine the product to investigate possible problems that can be experienced by users (Schmettow, 2012). The number of users required for usability studies has been a debatable issue for more than 30 years (Lewis, 1982; Schmettow, 2012). Specific numbers have already been suggested but these numbers are typically related to the studies which attempt to identify usability problems of a particular product (Eraslan et al., 2016a). The most popular example is from Nielsen and Landauer (1993) who suggest that it is possible to discover 85% of usability problems of a particular product with only five users. In contrast, Faulkner (2003) suggests that this is not valid for all five users. Based on her study, 15 users are required instead of five users. Another example is from Hwang and Salvendy (2010) who analysed many published research papers since 1990 and suggest that at least 10±2 users are needed to discover 80% of usability problems. However, there is also a considerable number of researchers who argue that each study has its own features, thus the number of users cannot be unique for all studies (Alroobaea & Mayhew, 2014; Bevan et al., 2003; Cazañas, de San Miguel, & Parra, 2017; Schmettow, 2012). For instance, the complexity of products and/or the characteristics of users can cause differences in the required number of users (Caulton, 2001; J. Spool & Schroeder, 2001).

Existing research in this field does not focus on eye tracking studies which are commonly used to evaluate and improve usability of web pages (Ehmke & Wilson, 2007). Since eye tracking studies are usually time-consuming, it is crucial for researchers to estimate the ideal sample size for their studies. Pernice and Nielsen (2009) suggest to have 39 users for a stable heat map to highlight which parts of web pages get more attention. However, as shown in Figure 2, these maps do not show eye movement sequences and they are much easier to analyse in comparison with scanpaths. Therefore, we should not assume that this number is also appropriate for scanpath analysis. Moreover, Pernice and Nielsen (2009)

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

suggest that a qualitative study can be conducted with six users and then their eye movements can be watched. Although researchers may gain valuable insights for determining which fixation features (such as, fixation duration, fixation count, time to first fixation, etc.) would be relevant for their further analyses by watching eye movements of users, the details are likely to be lost as an eye tracking dataset typically consists of many fixations. Therefore, detailed data analysis is crucial and becomes critical compared to qualitative analysis. Furthermore, eye tracking data analysis also needs to be related to certain behavioural tasks. Otherwise, the analysis cannot be completed to understand the user behaviour.
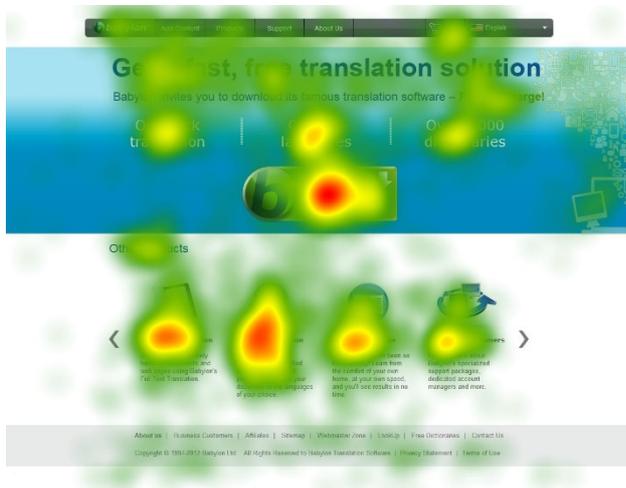

Figure 2. A heat map on the Babylon web page

The effects of the number of users on scanpath analysis were investigated by using a scanpath analysis algorithm called Scanpath Trend Analysis (STA) on six web pages segmented with the VIPS approach (Eraslan et al., 2016a). This algorithm discovers the trending scanpath on a particular page in terms of its AOIs (Eraslan et al., 2016b). This path is different from an absolute path that is shared by all users. Other existing approaches typically try to discover an absolute path (Eraslan, Yesilada, & Harper, 2015). However, their resulting paths tend to be very short because of the variations caused by individual differences, and therefore they typically have low similarities to the user scanpaths (Eraslan et al., 2015, 2016b). In comparison with other approaches, the STA algorithm provides the path with the highest similarities to the user scanpaths, thus it would be more helpful for behaviour analysis (Eraslan et al., 2016b). The details of the STA algorithm is provided in the following section.

The study with the STA algorithm showed that it is possible to achieve almost the same results with a less number of users (Eraslan et al., 2016a). Specifically, there is a possibility of achieving 75% similarity to the results of 65 users with 27 users for searching tasks and 34 users for browsing tasks. However, the study did not show whether this possibility is valid with other page segmentation approaches. Thus, we ask whether it matters how we segment a web page.

Methodology

This section firstly gives a brief information about the STA algorithm, secondly explains our eye tracking study, and then describes the segmentation approaches used and our analysis procedure.

STA: Scanpath Trend Analysis

The STA algorithm has the following three stages: (1) Preliminary Stage, (2) First Pass, and (3) Second Pass. These stages are briefly described below. The full description of the STA algorithm can be found in (Eraslan et al., 2016b).

Preliminary Stage. The algorithm initially takes a series of fixations for each user on a web page and the AOIs of the page. It then finds the corresponding AOI for each fixation to represent the user scanpaths in terms of the AOIs. For example, the scanpath in Figure 1 is represented as MPPQ as the user fixated M, P, again P and Q respectively. The durations of these fixations are also stored.

First Pass. The aim of the first pass is to analyse the user scanpaths for the discovery of the trending AOIs based on the total number of fixations and the total duration of fixations (dwell time) on the AOIs. A user can fixate the same AOI more than once consecutively (such as, M**PP**Q) and/or non-consecutively (such as, M**P**Q**P**). We refer to each non-consecutive visit as an instance. For example, there are two instances of P in M**P**Q**P**. As the STA algorithm performs sequential analysis, these instances should be differentiated. To differentiate the instances of the same AOI, different numbers are assigned to them where the longest instance gets the first number. For example, if a user fixates M, M, P, M and M for 100 ms, 200 ms, 200 ms, 400 ms and 200 ms respectively, his/her scanpath is represented as M2 [100 ms] M2 [200

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

ms] P1 [200 ms] M1 [400 ms] M1 [200 ms] because the dwell time of the first instance of M (100 ms + 200 ms) is less than the dwell time of its second instance (400 ms + 200 ms).

When the algorithm differentiates the instances, it starts the discovery of the trending instances. When the total number of fixations on a particular instance is greater or equal to the minimum total number of fixations on the fully shared instances of the user scanpaths, and the dwell time on the instance is greater or equal to the minimum dwell time on the fully shared instances, the instance is defined as a trending instance. After the discovery of the trending instances, the algorithm removes other instances from the user scanpaths as they will not be in the trending path.

Second Pass. In the second pass, the user scanpaths are firstly collapsed by merging the same instances to determine the exact positions of the instances in the scanpaths because the second pass computes the sequential priority value for each instance in each scanpath by using the positions. For example, when the following scanpath is available M1 [150 ms] P1 [125 ms] P1 [150 ms] Q1 [125 ms], the positions of M1, P1 and Q1 should be zero, one and two respectively (the starting position is zero). When the algorithm merges the same instances, it combines them into one instance where its duration is equal to the total duration of the instances (such as, M1 [150 ms] P1 [125 ms] P1 [150 ms] Q1 [125 ms] → M1 [150 ms] P1 [275 ms] Q1 [125 ms]). The total number of fixations on each instance (i.e., the number of occurrence for each instance) is also stored. After that, the sequential priority value (ψ) of each instance in each user scanpath is computed with Equation (1) where P represents the instance position in the scanpath and L represents the length of the user scanpath. 1 and 0.1 are given the *max* and *min* values respectively to give 1 point the first instance and 0.1 points to the last instance.

$$\psi = 1 - P \cdot \frac{max - min}{L - 1} \tag{1}$$

The total priority value (Ψ) for each instance is then computed with Equation (2) where n represents the number of user scanpaths. In particular, if there are five different instances in total (such as, M1, M2, P1, P2 and Q1), then the total priority value is calculated for each of them. The total priority values of the instances are then used by the algorithm to locate the instances in the trend-

ing path in descending order, thus the overall positions of the instances in the user scanpaths are preserved. If more than one instance share the same total priority value, their dwell time and the total number of fixations on the instances are also taken into consideration. In the end, the algorithm removes the numbers of the instances (such as, M1 → M), and then deletes the consecutive repetitions (such as, MPPQ → MPQ) to represent the trending scanpath in terms of the AOIs.

$$\Psi = \sum_{i=1}^{n} \psi_i \tag{2}$$

As an example, Figure 3 illustrates the trending scanpath on the Babylon page which was constructed with the STA algorithm by using the scanpaths in Figure 4. The trending scanpath shows us which AOIs are most commonly used and in which order. If we remove the unused AOIs and re-organise the remaining ones based on the trending scanpath, we would provide faster download of the page and direct access to the commonly used AOIs based on the desired order.
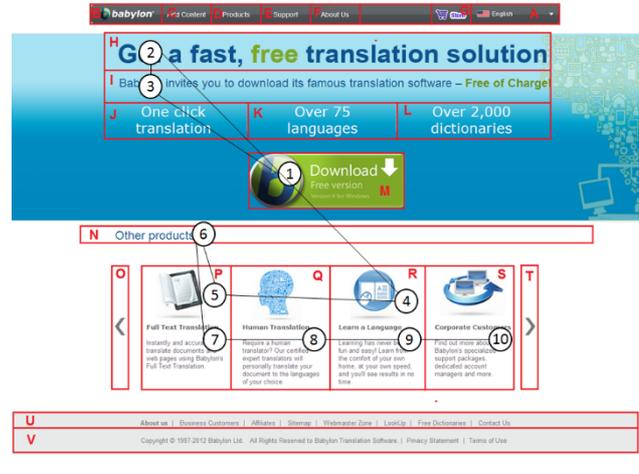
Figure 3. The trending scanpath on the Babylon web page

Eye Tracking Study

To investigate whether the possibility of achieving almost the same results with fewer users in scanpath trend analysis (Eraslan et al., 2016a) is restricted to a particular segmentation approach, we used the same dataset which is from an eye tracking study. This study is briefly described below and its full description can be found in (Eraslan et al., 2016a, 2016b).
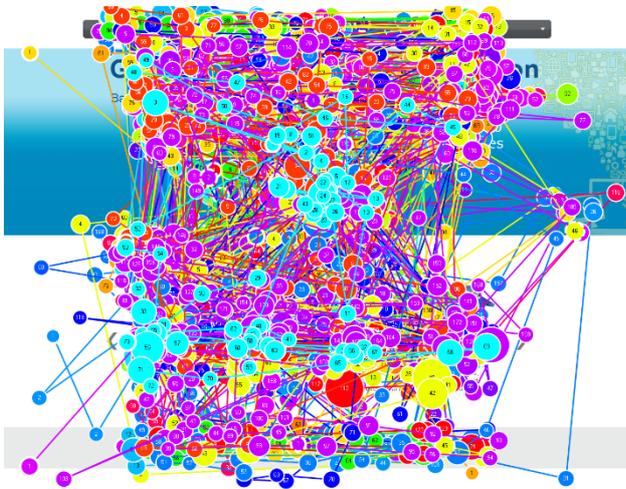
Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Figure 4. Multiple scanpaths on the Babylon web page

Equipment. Tobii T60 eye tracker was used to record the eye movements of the users. It was built-in a 17" monitor and the screen resolution of the monitor was adjusted to 1280 x 1024.

Pages. The web pages used in the eye tracking study were randomly chosen from a group of pages used by Akpınar and Yeşilada (2013). In their study, they investigated the visual complexities of the home pages of the top 100 websites listed by Alexa.com by using the ViCRAM tool (Michailidou, 2006), and then created three groups of pages based on their complexity (low, medium and high) where each group contains 10 randomly chosen pages. For this eye tracking study, two pages were randomly selected from each group: Apple & Babylon from the low complexity group, AVG & Yahoo from the medium complexity group, Godaddy & BBC from the high complexity group (see Appendix).

Some of the web pages contain much more text compared to others. For example, the Yahoo page contains much more text (353 words) compared to the Apple page (86 words). To quantitatively illustrate the text density of the pages, we referred to the text density function in Equation (3) given by J. M. Spool, Scanlon, Snyder, Carolyn, and DeAngelo (1999). Since the pages were shown in a fixed screen, we only report the total number of words seen by the participants on each page in Table 1.

$$Text\ Density = \frac{Total\ number\ of\ words\ on\ a\ page}{Length\ of\ printed\ page\ in\ inches} \quad (3)$$

Table 1. The total number of words on a page

Page	The total number of words on a page
Apple	86
Babylon	156
AVG	162
Yahoo	353
Godaddy	163
BBC	300

User Tasks. The participants were requested to perform two different kinds of tasks on the web pages called browsing and searching tasks. For the browsing tasks, there was no specific objective, so the participants freely browsed on the pages. For the searching tasks, the participants needed to locate some specific information or items on the web pages. The searching tasks used in the eye tracking study are provided in Table 2.

Table 2. The searching tasks used in the eye tracking study

Page	Tasks
Apple	(a) Can you locate a link which allows watching the TV ads relating to iPad mini?
	(b) Can you locate a link labelled iPad on the main menu?
Babylon	(a) Can you locate a link that you can download the free version of Babylon?
	(b) Can you find and read the names of other products of Babylon?
AVG	(a) Can you locate a link which you can download a free trial of AVG Internet Security 2013?
	(b) Can you locate a link which allows you to download AVG Anti-virus Free 2013?
Yahoo	(a) Can you read the titles of the main headlines which have smaller images?
	(b) Can you read the first item under the News title?
Godaddy	(a) Can you find a telephone number for technical support and read it?
	(b) Can you locate a text box where you can search for a new domain?
BBC	(a) Can you read the first item of the Sport News?
	(b) Can you locate the table that shows market data under the Business title?

Procedure. The study was performed in a quiet room. Before the participants viewed the web pages, they read an information sheet about the study and singed a consent form. Their gender, age-group and education level were then asked. The participants were also asked to rank the web pages based on their usage (Daily, Weekly, Monthly, Less than once a month, Never). After that, they viewed the web pages twice in a random order with counterbalancing for the browsing (30 seconds) and searching tasks (max 120 seconds). For example, one participant completed the tasks in the following order: "Babylon - Search, AVG - Search, Apple - Browse, Godaddy - Search Yahoo - Browse, Babylon - Browse, Apple - Search, Yahoo - Search, Godaddy - Browse, AVG - Browse, BBC -

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Browse and BBC – Search" and the other one completed in this order: "BBC - Search, Babylon - Browse, BBC - Browse, Babylon - Search, AVG - Search, Godaddy - Browse, AVG - Browse, Apple - Search, Yahoo - Browse, Yahoo - Search, Apple - Browse, Godaddy - Search'". The participants could be familiar with the pages during their first visits and this situation could affect how they interact with the pages during their second visits. Hence, for each web page, one half of the participants firstly completed the searching task and then the browsing task whereas another half firstly completed the browsing task and then the searching task. The tasks were read by the researcher to the participants when the relevant pages were shown on the screen during the eye tracking sessions. The participants were not allowed to use a keyboard and a mouse as the tasks could be completed by only scanning the pages. When they completed their eye tracking sessions, they were asked to draw what they remember on the pages. Apart from the fixations made by the participants after completing the searching tasks, none of the fixations were excluded from the analysis.

Participants. The eye tracking study conducted with 81 users at two different universities with 40 female and 41 male users. Most of these users were students from the universities and they were mainly between the ages of 18-34. The pages used in this study were not regularly visited by the users. The percentage of the users who reported that they never visited the pages or visited them less than once a month were as follows: Apple: 71.6%, Babylon: 93.8%, AVG: 92.6%, Yahoo: 61.7%, Godaddy: 98.8% and BBC: 46.9%.

Page Segmentation Approaches

Web pages can be segmented automatically or manually based on the goals of studies (Yesilada, 2011). In the previous study to investigate the possibility of achieving almost the same results with less users in scanpath trend analysis, the experiments were conducted with the extended and improved version of the VIPS approach to automatically segment the pages (Akpınar & Yeşilada, 2013; Eraslan et al., 2016a). In this current paper, we used the BOM approach for the automatic segmentation of the web pages (Sanoja & Gançarski, 2014). Similar to the VIPS approach, the BOM approach also uses the DOM structure and the visual representation of web pages but it computationally applies different algorithms. Specifically, in the BOM approach, the segments are also grouped based on the four Gestalt laws (Proximity, Similarity,

Closure and Simplicity). Both of the segmentation approaches have also a granularity level parameter which affects the size of segments and the number of segments. In our experiments, we used the BOM approach with its default value (0.3) for this parameter.

The current publicly available implementation of the BOM approach (http://www-poleia.lip6.fr/~sanojaa/BOM/) could not properly segment some of the web pages. Thus, we followed a systematic approach to fix the problems (Eraslan, Yesilada, & Harper, 2016c) by considering the algorithm given in their papers (Sanoja & Gançarski, 2013, 2014): (1) If an element covered only space, the element was ignored. (2) If a larger element covered other smaller elements and not all parts of this larger element were covered by the smaller elements, the larger element was used. This was to ensure that all parts of the larger element are covered by an element. (3) If there were overlaps between multiple elements, the borders of the elements were adjusted to deal with the overlaps. (4) If there was an element which was not located in any element as a result of the segmentation, it was located in the nearest element. As an example, Figure 5 shows how the Babylon page segmented by the current publicly available implementation of the BOM approach. The menu item for the language selection was not covered by any block, and therefore it was located in its nearest block L13 (The menu item "Store"). Figure 6 shows how the Babylon page was divided into their elements by using the BOM approach with a systematic approach for fixing the problems.

We also followed a systematic way to find other automatic page segmentation approaches with a public implementation by looking at the relevant papers and searching from the web. As illustrated in Table 3, other segmentation approaches do not have their public implementations and/or they are not applicable (NA) for this study. If we could not find the implementation of the proposed approach, we sent an email to the author(s) to ask whether they could provide us the implementation of their proposed segmentation approach. If the authors did not respond to our email, we assume that their segmentation approaches do not have a public implementation. Moreover, not all segmentation approaches are suitable for our study. In particular, if a segmentation approach focusses on getting specific parts of web pages, it is considered as not applicable (NA) for our study as we are not interested in only specific parts of web pages. For exam-

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

ple, one approach focusses on getting the main content of a web page instead of segmenting the web page into its visual elements (Kohlschütter, Fankhauser, & Nejdl, 2010). There are also some approaches which are outdated such as the approach of Milic-Frayling and Sommerer (2002) which focusses on HTML table tags. These approaches are also classified as NA. Although the page segmentation approach of Michailidou, Harper, and Bechhofer (2008) is publicly available, it has some bugs and does not generate a proper segmentation.

Figure 5. The Babylon page segmented by the current publicly available implementation of the BOM approach
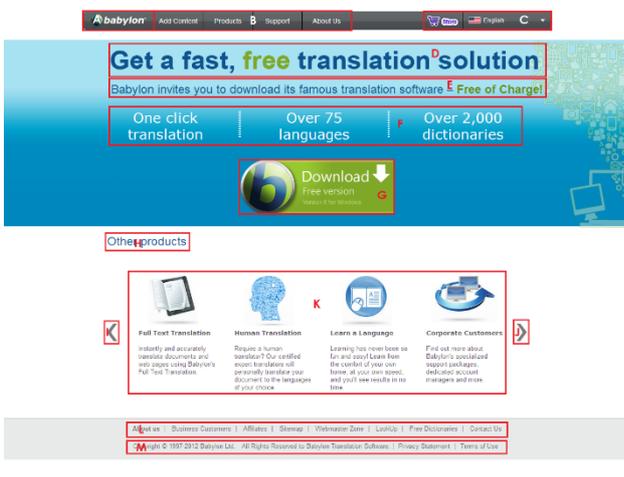

Figure 6. The Babylon page segmented by the BOM segmentation approach [The segmentation problems highlighted in Figure 5 are addressed in this version]

Table 3. The public availability of the page segmentation approaches [✓: Available, ✗: Not Available, NA: Not Applicable] - The references are in (Yesilada, 2011)

Category	Source	Availability
Clustering	Alcic and Conrad (2011)	✗
Custom	Cheng and Gotz (2009)	NA
	Yang and Shi (2009) & Xiang et al. (2007)	✗
	Guo et al. (2007)	✗
	Gupta et al. (2007)	NA
	Hattori et al. (2007)	✗
	Gu et al. (2002)	✗
	Chen et al. (2001)	NA
	Sanoja and Gançarski (2014) & Sanoja and Gançarski (2013) - BOM	✓
	Wei, Lu, Li, and Liu (2015)	✗
	Zeleny, Burget, and Zendulka (2017)	✗
DOM-Based	Fauzi et al. (2009)	NA
	Vineel (2009)	✗
	Xiao et al. (2008)	✗
	Chen et al. (2005, 2003)	✗
	Yin and Lee (2005)	✗
	Liu et al. (2004)	✗
	Lin and Ho (2002)	NA
Heuristics	Toh and Hong (2014)	✗
	Burget and Rudolfova (2009) Burget (2007)	NA
	Ahmadi and Kong (2008)	✗
	Michailidou et al. (2008)	✓
	Kreuzer, Hage, and Feelders (2015)	NA
	Debnath et al. (2005)	✗
	Kovacevic et al. (2002)	✗
	Milic-Frayling and Sommerer (2002)	NA
Image Processing Algorithms	Cao et al. (2010)	✗
Machine Learning	Chakrabarti et al. (2008)	✗
	Borodin et al. (2007)	NA
	Mahmud et al. (2007)	NA
	Baluja (2006)	✗
	Bing, Guo, Lam, Niu, and Wang (2014)	NA
	Feng, Zhang, Wu, and Wang (2016)	✗
Pattern Matching	Xiang and Shi (2006)	✗
	Nanno et al. (2004)	✗
	Cuzzolaa, Jovanović, Bagheri, and Gašević (2015)	✗
	Whang et al. (2001) & Hwang et al. (2003)	NA
Ranking-Based	Yin and Lee (2004)	NA
Text-Based	Sun et al. (2011)	NA
	Kohlschütter et al. (2010)	NA
	Kohlschütter (2009)	NA
	Kohlschütter and Nejdl (2008)	✗

We also manually segmented the web pages as manual segmentation can also be preferred by researchers based on their goals (Owens et al., 2011). In order to be more objective, we did not want to do the segmentation

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

by ourselves, and therefore we used the recall approach. At the end of the eye tracking study described above, the participants were asked to draw what they remember about the layouts of the web pages on a piece of paper after they completed their tasks on the web pages. Figure 7 and Figure 8 shows how the Babylon web page was drawn by two participants. We systematically combined their drawings to identify the AOIs of the web pages. To be more objective, three researchers firstly combined the drawings for each web page, including other researchers who are not authors of this paper. For example, they noted how many users divided the main menu into its elements and how many of users drew the main menu as an entire element. Once they combined the drawings, their notes were analysed to see all the elements drawn by the users. If there was a larger element that covers other smaller elements, the smaller elements were used as they were more specific. For example, if a menu was divided into their items, the menu items were considered as separate elements instead of using the menu as one element. Even though we applied the same approach here as explained for the BOM approach (Sanoja & Gançarski, 2014) (If a larger element covered other smaller elements and not all parts of this larger element were covered by the smaller elements, the larger element was used), we used the smaller elements since all parts of the larger element were covered by the smaller elements here. As an example, Figure 9 shows how the Babylon page is segmented by using this approach.
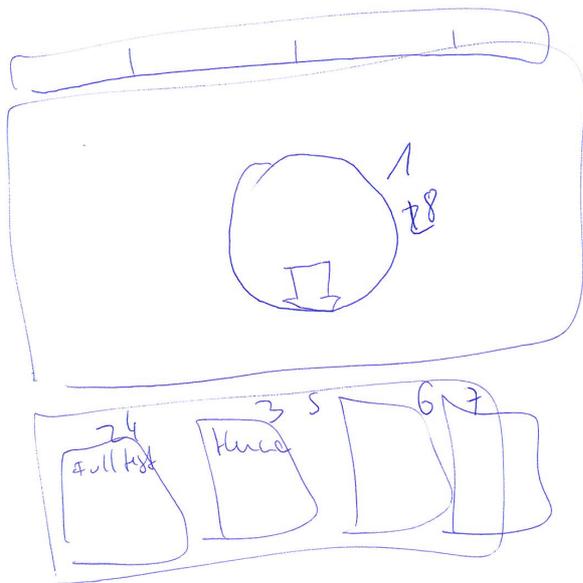

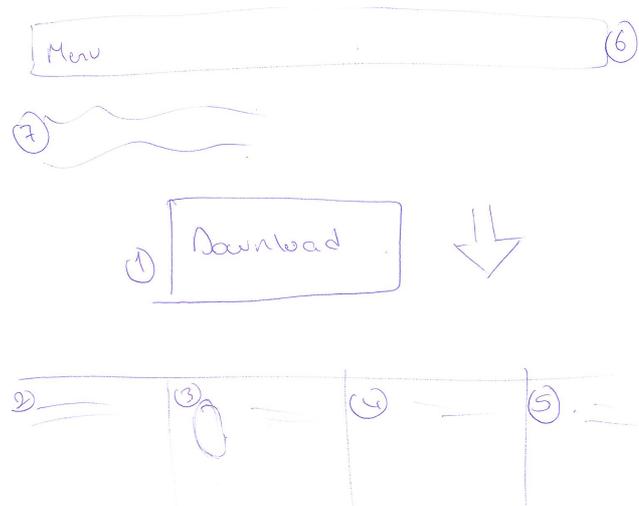Figure 7. The Babylon page drawn by the participant 1

Figure 8. The Babylon page drawn by the participant 2

Figure 9. The Babylon page segmented by the user driven segmentation approach

The overview of the differences between the segmentations of the VIPS, BOM and user driven approaches in terms of the number of AOIs generated and their coverage on the web pages can be seen from Table 4. Some parts of a particular web page may be located in separate AOIs with some segmentation approaches whereas the same parts may be located in only one AOI with other approaches. For example, each of the other Babylon products ("Full Text Translation", "Human Translation", "Learn a Language" and "Corporate Customers") was located in a separate AOI with the VIPS (with its most satisfactory granularity (Akpınar & Yeşilada, 2013)) and user driven segmentation approaches (see Figure 1 and

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Figure 9) however all these products are located in the same AOI with the BOM segmentation approach (with its default granularity parameter) (see Figure 6). As illustrated in Table 4, the VIPS approach provides more and smaller segments on the Babylon, AVG, and BBC web pages in comparison with the BOM and user driven approaches. In contrast, it provides fewer and larger segments on the Yahoo page.

Table 4. The % of the coverage of the AOIs generated by the VIPS, BOM and user driven approaches [N: The number of AOIs, M: Mean, MD: Median, SD: Standard Deviation]

Page	Approach	N	M	MD	SD	Max	Min
Apple	VIPS	18	2.68	0.82	5.46	23.57	0.06
	BOM	7	7.54	3.38	9.57	26.58	0.77
	User driven	18	2.58	0.76	4.84	21.03	0.30
Babylon	VIPS	22	1.98	1.60	1.65	5.27	0.24
	BOM	13	2.79	1.05	4.39	16.09	0.06
	User driven	17	2.62	0.56	3.98	15.58	0.14
AVG	VIPS	25	1.97	0.45	4.65	20.93	0.07
	BOM	7	6.99	1.82	8.16	21.05	0.32
	User driven	20	2.41	2.47	1.90	8.35	0.44
Yahoo	VIPS	10	6.93	1.94	9.57	25.78	0.19
	BOM	12	5.09	1.56	7.13	24.78	0.08
	User driven	24	2.57	0.92	2.88	9.14	0.10
Godaddy	VIPS	16	4.16	0.52	9.44	35.15	0.10
	BOM	11	5.64	1.46	7.21	18.42	0.11
	User driven	19	2.94	1.18	4.36	17.61	0.35
BBC	VIPS	21	3.22	0.95	4.26	12.90	0.15
	BOM	5	14.43	8.79	16.28	43.22	4.24
	User driven	10	6.72	6.62	3.85	14.45	0.17

The segmentation of the pages with each of these segmentation approaches is visualised and provided in our external repository. Once the web pages were segmented by using both the BOM segmentation approach (Sanoja & Gançarski, 2014) and the user driven approach, we replicated the experiments of our previous study (Eraslan et al., 2016a) by using the same methodology as described in the following section.

Analysis Procedure

As some of the users could not successfully complete their tasks on some of the web pages and/or experienced some eye calibration problems, they were classified as unsuccessful users and excluded from the experiments. If a particular user was unsuccessful on the page X but successful on the page Y, s/he was excluded from only the page X. Due to this reason, the number of users on the

web pages for the browsing and searching tasks were not the same. However, it would be better to have the same number of users on all the pages for both types of the tasks for the consistency among the pages and estimation of the overall graph. Thus, some of the successful users were also randomly excluded from the experiments. As a result, the experiments were conducted with 65 users on each web page for the browsing and searching tasks.

The STA algorithm was then applied to the users to discover their trending scanpaths on the pages which were segmented by using the BOM approach (Sanoja & Gançarski, 2014) and the user driven approach. Following this, smaller groups of users were chosen with the size of one user to 64 users from the successful users and then the STA algorithm was applied to these groups to discover their trending scanpaths. In other words, sub-groups were randomly created with different sizes from 1 user to 64 users and their trending scanpath were identified with the STA algorithm (i.e., group of one user, group of two users, group of three users … group of 64 users). The results of the smaller groups were then compared with the results of the entire group. To deal with any possible effects of the selected users, 100 different combinations were generated for each group size and their median similarity was taken as a similarity to the entire the group to counteract extreme cases (Faulkner, 2003).

To compare two scanpaths, the Levenshtein Distance (String-edit) algorithm was used as it has widely been used to determine the distance between two scanpaths (Josephson & Holmes, 2002; Takeuchi & Habuchi, 2007). This algorithm transforms one scanpath to another by using the minimum number of addition, deletion and substitution operations. The minimum number of operations shows the distance between the two scanpaths. For instance, the distance between DEFG and DEHG is equal to one because it is sufficient to substitute F with H to transform one of them to another or vice versa.

The sizes of AOIs and the distances between the AOIs are potentially different, as a consequence the substitution costs between all pairs of AOIs may not be the same (Josephson & Holmes, 2002). A substitution matrix has been suggested to take this issue into account (Takeuchi & Habuchi, 2007). This matrix should include substitution costs between all pairs of AOIs. As suggested by Takeuchi and Habuchi (2007), the Euclidean distances between the AOIs were used to generate a substitution matrix in the experiments. Equation (4) below illustrates

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

how to calculate a substitution cost between two AOIs I and J based on the Euclidean distance where I_1 and I_2 are x and y coordinates of the centre of the element I and α is a type of normalisation parameter (Takeuchi & Habuchi, 2007). The normalisation parameter was taken as 0.001 (Takeuchi & Habuchi, 2007). The matrix was then integrated into the String-edit algorithm to determine the distance between two scanpaths by computing the minimum cost for transforming one scanpath to another.

$$f(\vec{I}, \vec{J}) = \alpha \sum_{i=1}^{2} \sqrt{(I_i - J_i)^2} \tag{4}$$

To find a similarity between two scanpaths as a percentage based on the String-edit distance (Underwood, Humphrey, & Foulsham, 2008), the distance (D) between two scanpaths was firstly divided by the length of the longer scanpath (L) to calculate a normalised score to prevent possible inconsistencies that can be caused by different lengths. The normalised score was secondly subtracted from one and finally multiplied by 100. The relevant formula is shown in Equation (5). For example, the similarity between DEFG and DEHG is calculated as 75% as the distance (D) between these scanpaths is equal to one (the substitution between F and H), and the length of the longer scanpath (L) is equal to 4 (both of them have the same length in this example).

$$Similarity = 100 \cdot (1 - \frac{D}{L}) \tag{5}$$

As we investigated how many users are needed to achieve almost the same results as 65 users, our dependent variable was the number of users and our independent variable was the similarity to the results of 65 users. Once we calculated the similarity to the entire group for each smaller group, the curve estimation feature of SPSS (http://www.ibm.com/analytics/us/en/technology/spss/) was applied to select the best curve that fits into the mean of these similarities on the six web pages for the browsing and searching tasks. Specifically, the curve with the minimum standard error of the estimate was selected.

Results

Figure 10 and Figure 11 show the estimated curves by using the VIPS (Akpınar & Yeşilada, 2013), BOM (Sanoja & Gançarski, 2014) and user driven segmentation

approaches on the six web pages for the browsing and searching tasks with their equations where S is the similarity to the entire group and i is the number of users (Eraslan et al., 2016a). The standard errors of the estimated curves for the VIPS, BOM and user driven approaches are 0.073, 0.066 and 0.034 for the searching tasks, and 0.044, 0.063 and 0.082 for the browsing tasks respectively.

To compare two estimated curves, the absolute difference in the curves was calculated for each group size as illustrated in Table 5. For instance, the difference between the curves generated by using the VIPS approach and the BOM approach for three users is equal to 5.87 (With the BOM: 19.97, With the VIPS: 25.84) for the browsing tasks and 4.88 (With the BOM: 23.73, With the VIPS: 28.61) for the searching tasks.

Table 5. The differences between the curves generated by using the VIPS approach and the BOM approach where i is the number of users.

Task	i	With the BOM	With the VIPS	Difference
Browsing	1	11.91	15.94	4.02
	2	16.50	21.62	5.12
	3	19.97	25.84	5.87
	4	22.86	29.33	6.47
	5	25.38	32.35	6.97
	…	…	…	…
Searching	1	13.79	17.67	3.87
	2	19.42	23.95	4.52
	3	23.73	28.61	4.88
	4	27.36	32.47	5.11
	5	30.54	35.81	5.26
	…	…	…	…

Table 6 also illustrates the mean, median, maximum, minimum and standard deviation (SD) of the differences between the curves generated by using the VIPS, BOM and user driven approaches. As illustrated in the table, the mean difference between the curves generated by using the VIPS approach and the BOM approach is only 4.13% (SD: 1.13%) for the searching tasks and 11.90% (SD: 2.76%) for the browsing tasks. Furthermore, the mean difference between the curves generated by using the VIPS approach and the user driven approach is only 5.88% (SD: 1.47%) for the searching tasks and 11.49% (SD: 1.62%) for the browsing tasks. Finally, the mean difference between the curves generated by using the BOM approach and the user driven approach is only 2.50% (SD: 1.68%) for the searching tasks and 3.23% (SD: 1.83%) for the browsing tasks.

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis
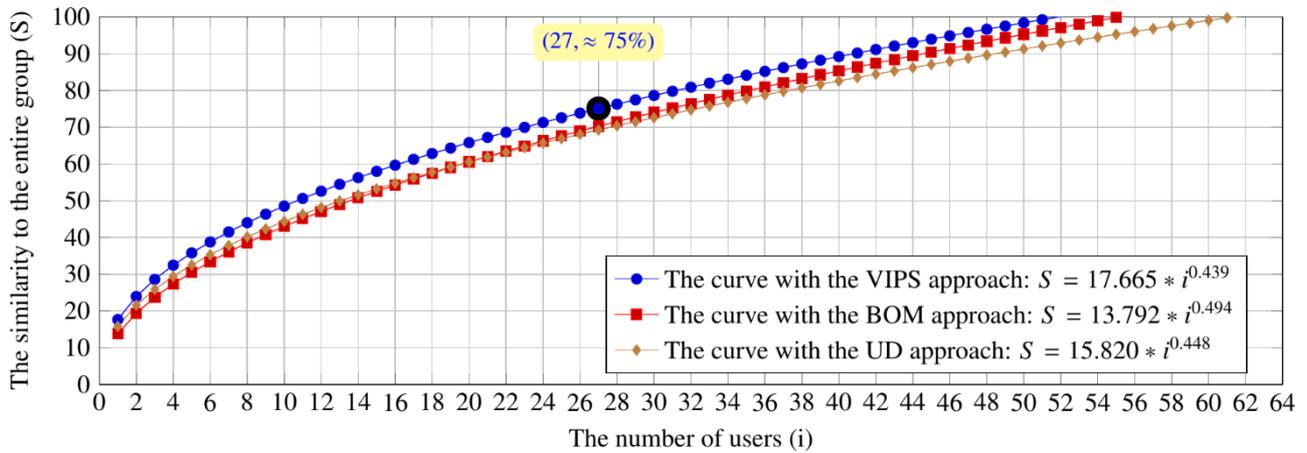
Figure 10. The estimated curves by using the VIPS, BOM and user driven segmentation approaches on the six web pages for the searching tasks
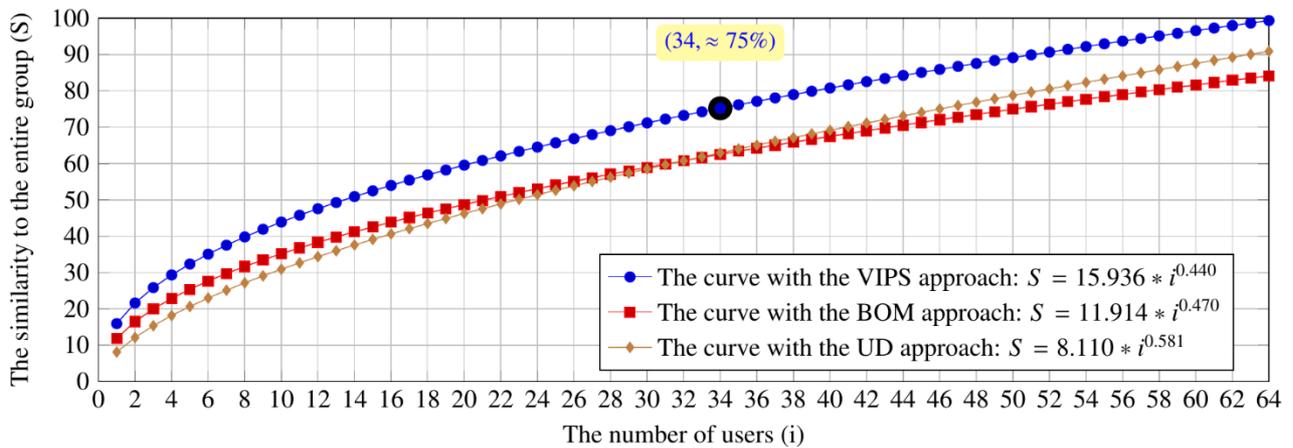

Figure 11. The estimated curves by using the VIPS, BOM and user driven segmentation approaches on the six web pages for the browsing tasks

To sum up, the previous and current results are mainly consistent with ~5\% difference for the searching tasks and ~10\% difference for the browsing tasks.

Table 6. The mean (M), median (MD), maximum, minimum and standard deviations (SD) of the differences between the curves generated by using the VIPS approach, the BOM approach and the user driven approach where i is the number of users

Task	Approaches	M	MD	Max	Min	SD
Browse	VIPS-BOM	11.90	12.60	15.27	4.02	2.76
	VIPS-User driven	11.49	11.85	13.37	7.83	1.62
	BOM-User driven	3.23	3.39	6.94	0.00	1.83
Search	VIPS-BOM	4.13	4.33	5.53	1.96	1.13
	VIPS-User driven	5.88	6.22	7.75	1.85	1.47
	BOM-User driven	2.50	2.08	5.79	0.04	1.68

Discussion

Based on the results presented in this paper, we can suggest that we can see almost the same trend in the curves when we use the VIPS (Akpınar & Yeşilada, 2013), BOM (Sanoja & Gançarski, 2014) and user driven approaches to segment the web pages. In particular, the results obtained from the BOM and the user driven approaches are very close to each other. Therefore, we can suggest that the previous and current results are consistent with some small deviations (~5\% difference for the searching tasks and ~10\% difference for the browsing tasks).

Although the VIPS approach is popular among researchers to segment web pages, they can also use other

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

page segmentation approaches. However, the previous study did not show whether the possibility of achieving almost the same results with fewer users is limited to the VIPS approach. Therefore, when researchers want to use another segmentation approach with the STA algorithm, they could not ensure whether the possibility is still valid for them. Hence, this study is beneficial for them from this aspect as it shows that we can approximate almost the same results with a smaller group of users in scanpath analysis regardless of the segmentation approach used.

As mentioned in our related work section, although other aspects of eye tracking data (especially, heat maps (Pernice & Nielsen, 2009)) have been studied, the number of users required for scanpath analysis has not been studied in depth. The possibility of achieving almost the same results with less users in scanpath trend analysis has been raised but the segmentation effect on this possibility was not investigated. Therefore, this study is an important step forward in existing research. Furthermore, the results are still consistent with the suggestions for heat maps (Pernice & Nielsen, 2009), especially for the browsing tasks, even though the analysis of heat maps are much easier in comparison with scanpath analysis (see Figure 2 and Figure 4).

We also observed some differences in the curves generated by using the VIPS, BOM and user driven segmentation approaches. Specifically, the VIPS approach has higher similarities compared to other two segmentation approaches, especially for the browsing tasks. The differences in the curves generated for the browsing tasks are higher in comparison with the differences in the curves generated for the searching tasks. Since the participants did not require to locate specific items on the web pages for the browsing tasks and they were allowed to browse freely, the variations between the user scanpaths tend to be higher, and therefore it is quite normal to see higher difference in the curves generated for the browsing tasks. We think that the differences in both the searching and browsing tasks might be caused by the sizes of the AOIs. In particular, when the BOM approach is used to automatically segment the web pages, fewer and larger AOIs are usually generated. Therefore, user scanpaths become shorter, as a consequence their resulting paths also become shorter. In these cases, small deviations between two scanpaths considerably decrease the similarity between the scanpaths. For example, the similarity between the scanpaths ABC and ACB is calculated as 33.33% by

the standard String-edit algorithm, even though they have the same elements (Josephson & Holmes, 2002). To investigate this expectation, we should conduct further experiments as we cannot say that the VIPS approach creates smaller blocks in all the cases in comparison with other approaches. In the further experiments, the same methodology can be applied by segmenting web pages with different granularity levels and the effects of the AOI sizes can be investigated. The reason for the VIPS approach having higher similarities could also be that it might provide the most similar AOIs to the AOIs which were actually used by the participants. However, this possible reason should be further investigated by determining the similarities between the AOIs generated by the segmentation approaches and the AOIs which were actually used by the users. However, in order to do this, a systematic approach should firstly be developed to discover the borders of the AOIs which were actually used. These borders can then be used for analysis

A study conducted by Akpınar and Yeşilada (2013) suggests that more and smaller segments are preferred by users instead of fewer and larger segments for the VIPS approach. Therefore, in the previous study (Eraslan et al., 2016a), the experiments was conducted by using the suggested granularity level of the VIPS algorithm. However, in our current experiments, we had to use the default granularity level of the BOM approach as no specific suggestion is provided. Since the granularity level affects the segmentation, it may also affect the results. Therefore, when a segmentation approach is proposed, its most successful/preferable/satisfactory granularity level should also be provided, if appropriate.

Because of the limited number of publicly available automatic segmentation approaches (Yesilada, 2011), we could only use the BOM approach for automatically segment the web pages to cross-check the previous findings obtained by using the VIPS approach. Since web pages can also be manually segmented by researchers based on their goals (Owens et al., 2011), we also applied a user driven approach to manually segment the web pages. However, further experiments need to be conducted by using more segmentation approaches and different eye tracking datasets with the same methodology to increase the generalizability of the probability of accessing almost the same results with fewer users. For example, it would be worthwhile to investigate how this possibility is affected when all users are familiar with web pages or when

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

web pages are segmented into very small or very large elements. Furthermore, a number of eye tracking experts could also be invited to discuss and identify the AOIs of the web pages based on their experience and then the same methodology could be applied. However, manual segmentation is typically based the goals of the study and it is almost impossible to estimate all possible segmentations. We could also apply the grid segmentation. However, the elements of the web pages could be inappropriately divided in that case as the web pages were not necessarily designed with a grid layout. For example, Figure 12 shows how the Babylon page could be segmented with a 4x4 grid segmentation. As illustrated, some elements are inappropriately divided, such as the text under the "Full Text Translation" title. Therefore, we did not use the grid segmentation in our study.
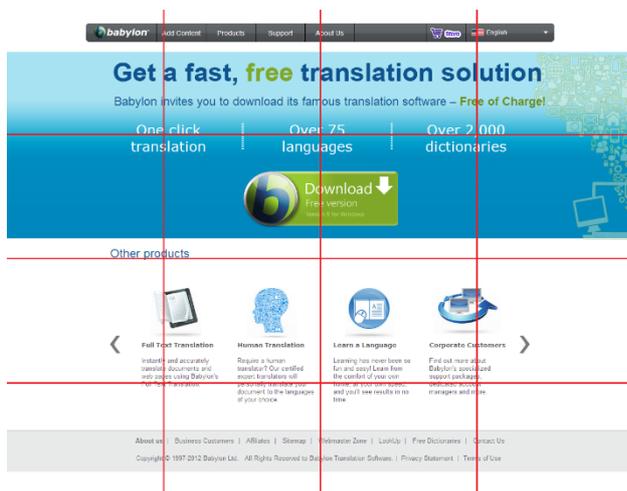

Figure 12. The Babylon page with a 4-4 grid segmentation.

In this study, only the STA algorithm was used but the methodology can also be applied to other scanpath analysis algorithms to investigate the effects of the number of users on their results. For example, the SPAM (Sequential Pattern Mining) algorithm was used by Hejmady and Narayanan (2012) to discover visual attention patterns during program debugging with an IDE (Integrated development environment). They conducted their study with 19 participants but they did not know how their results could be affected if they had less or more participants. As it is more reliable to report stable results, it is worthwhile to investigate the effects on the number of users on other existing scanpath analysis algorithms.

This study in not without limitations. The home page of six web pages were used. Although these web pages had different levels of visual complexities, more web pages with more complicated and different designs would be better to increase the generalizability of our findings. In this paper, we provided the results of the six web pages together as illustrated in Figures 10 and 11. However, the visual complexity and text density may affect the results. If we had more web pages in each visual complexity group (low, medium and high), we would be able to investigate whether the visual complexity of web pages affects the possibility of achieving almost the same results with a less number of users. We currently have two web pages in each visual complexity group. Thus, if we try to investigate the effects of the visual complexity, we will have only two pages in each visual complexity group and therefore the results would not be representative. Likewise, if we try to investigate the effects of the text density, we will have only one page with lower density (Apple: 86 words) and two pages with higher density (Yahoo: 353 word, BBC: 300 words), and therefore the results would not be conclusive. However, it would be interesting to investigate the effects of the visual complexity and text density on the possibility of achieving almost the same results with a less number of users. We are planning to explore these possible effects in the future studies. One can design a study to take into consideration both visual complexity and text density as features in selecting web pages

Most of the participants were university students and daily web users. Even though there were many participants with different backgrounds, the increase in the diversity of the users would be better to draw stronger conclusions. Furthermore, we only used the Levenshtein Distance algorithm with a substitution cost matrix to compute similarities between scanpaths. Albeit the Levenshtein Distance algorithm has widely been used in eye tracking research, other similarity measures, such as Needleman and Wunsch (1970), can also be used in the future and the findings can be cross-checked.

To sum up, this work is a step forward for better understanding of how the number of users affect scanpath analysis. It would benefit to both eye tracking and human computer interaction (HCI) researchers to estimate the ideal sample size for their studies by considering their time and budget when they want to conduct scanpath trend analysis. The methodology used is also promising for researchers who developed a scanpath analysis algo-

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

rithm and want to investigate the effects of the number of users on the results of their algorithms.

Conclusion

In this paper, we show that we can approximate almost the same results with a smaller group of users in scanpath trend analysis regardless of the segmentation approach used. The current findings concur with the findings of Eraslan et al. (2016a) suggesting that it is possible to achieve almost the same results with fewer users. Specifically, the findings of Eraslan et al. (2016a) suggest that it is possible to achieve 75% similarity to the results of 65 users with 27 users for searching tasks and 34 users for browsing tasks. In this paper, we investigated both the automatic and manual segmentation effects on the findings. Based on our experiments, we can suggest that our current findings are mainly consistent with the previous findings.

Although it might not be possible to provide exactly the same results with fewer users, 75% similarity is a promising value for practitioners because it means that approximately 3/4 of the results are the same. For example, the similarity between the scanpaths ABCDBADBDEFE and ABCEBADCDEF is calculated as 75% with the standard String-edit algorithm (Josephson & Holmes, 2002). The practical usefulness of 75% and other similarities will be investigated in the future. This paper allows practitioners to know more about possible differences when they have lower sample size, thus it helps them to determine their sample size based on their time and the budget of their studies.

Ethics and Conflict of Interest

The authors declare that the contents of the article are in agreement with the ethics described in http://biblio.unibe.ch/portale/elibrary/BOP/jemr/ethics.html and that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

References

Akpınar, M. E., & Yeşilada, Y. (2013). Vision Based Page Segmentation Algorithm: Extended and Perceived Success. In Q. Z. Sheng & J. Kjeldskov (Eds.), *Current Trends in Web Engineering: ICWE 2013 International Workshops ComposableWeb, QWE, MDWE, DMSSW, EMotions, CSE, SSN, and PhD Symposium, Aalborg, Denmark, July 8-12, 2013. Revised Selected Papers* (pp. 238-252). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-04244-2_22

Akpınar, M. E., & Yeşilada, Y. (2015). *"Old Habits Die Hard!": Eyetracking Based Experiential Transcoding: A Study with Mobile Users*. Paper presented at the Proceedings of the 12th Web for All Conference, New York, NY, USA. http://doi.acm.org/10.1145/2745555.2746646

Alroobaea, R., & Mayhew, P. J. (2014). *How many participants are really enough for usability studies?* Paper presented at the 2014 Science and Information Conference. http://doi.acm.org/10.1109/SAI.2014.6918171

Bevan, N., Barnum, C., Cockton, G., Nielsen, J., Spool, J., & Wixon, D. (2003). *The "Magic Number 5": Is It Enough for Web Testing?* Paper presented at the CHI '03 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA. http://doi.acm.org/10.1145/765891.765936

Bing, L., Guo, R., Lam, W., Niu, Z.-Y., & Wang, H. (2014). *Web Page Segmentation with Structured Prediction and Its Application in Web Page Classification*. Paper presented at the Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, New York, NY, USA. http://doi.acm.org/10.1145/2600428.2609630

Brown, A., Jay, C., & Harper, S. (2010). *Audio Access to Calendars*. Paper presented at the Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), New York, NY, USA. http://doi.acm.org/10.1145/1805986.1806028

Caulton, D. A. (2001). Relaxing the homogeneity assumption in usability testing. *Behaviour and Information Technology, 20*(1), 1-7. http://dx.doi.org/10.1080/01449290010020648

Cazañas, A., de San Miguel, A., & Parra, E. (2017). Estimating Sample Size for Usability Testing.

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Enfoqute(8), 172-185.
http://ingenieria.ute.edu.ec/enfoqueute/index.php/revista/article/view/126/133

Collberg, C., & Proebsting, T. A. (2016). Repeatability in Computer Systems Research. *Commun. ACM, 59*(3), 62-69. http://doi.acm.org/10.1145/2812803

Cuzzolaa, J., Jovanović, J., Bagheri, E., & Gašević, D. (2015). Automated Classification and Localization of Daily Deal Content from the Web. *Appl. Soft Comput., 31*(C), 241-256. http://dx.doi.org/10.1016/j.asoc.2015.02.029

Ehmke, C., & Wilson, S. (2007). *Identifying Web Usability Problems from Eye-tracking Data*. Paper presented at the Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1, Swinton, UK, UK. http://dl.acm.org/citation.cfm?id=1531294.1531311

Eraslan, S., Yesilada, Y., & Harper, S. (2015). Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research, 9*(1). https://bop.unibe.ch/index.php/JEMR/article/view/2430

Eraslan, S., Yesilada, Y., & Harper, S. (2016a). *Eye Tracking Scanpath Analysis on Web Pages: How Many Users?* Paper presented at the Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, New York, NY, USA. http://doi.acm.org/10.1145/2857491.2857519

Eraslan, S., Yesilada, Y., & Harper, S. (2016b). Scanpath Trend Analysis on Web Pages: Clustering Eye Tracking Scanpaths. *ACM Trans. Web, 10*(4), 20:21-20:35. http://doi.acm.org/10.1145/2970818

Eraslan, S., Yesilada, Y., & Harper, S. (2016c). *Trends in Eye Tracking Scanpaths: Segmentation Effect?* Paper presented at the Proceedings of the 27th ACM Conference on Hypertext and Social Media, New York, NY, USA. http://doi.acm.org/10.1145/2914586.2914591

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers, 35*(3), 379-383. https://doi.org/10.3758/BF03195514

Feng, H., Zhang, W., Wu, H., & Wang, C. J. (2016). *Web Page Segmentation and Its Application for Web Information Crawling*. Paper presented at the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). http://doi.acm.org/10.1109/ICTAI.2016.0097

Hejmady, P., & Narayanan, N. H. (2012). *Visual Attention Patterns During Program Debugging with an IDE*. Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, New York, NY, USA. http://doi.acm.org/10.1145/2168556.2168592

Hwang, W., & Salvendy, G. (2010). Number of People Required for Usability Evaluation: The 10±2 Rule. *Commun. ACM, 53*(5), 130-133. http://doi.acm.org/10.1145/1735223.1735255

Jay, C., Lunn, D., & Michailidou, E. (2008). End User Evaluations. In S. Harper & Y. Yesilada (Eds.), *Web Accessibility: A Foundation for Research* (pp. 107-126). London: Springer London. https://doi.org/10.1007/978-1-84800-050-6_8

Johnson, J. (2010). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 9780123750303

Josephson, S., & Holmes, M. E. (2002). *Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web*. Paper presented at the Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, New York, NY, USA. http://doi.acm.org/10.1145/507072.507081

Kohlschütter, C., Fankhauser, P., & Nejdl, W. (2010). *Boilerplate Detection Using Shallow Text Features*. Paper presented at the Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, NY, USA. http://doi.acm.org/10.1145/1718487.1718542

Kreuzer, R., Hage, J., & Feelders, A. (2015). A Quantitative Comparison of Semantic Web Page Segmentation Approaches. In P. Cimiano, F. Frasincar, G.-J. Houben, & D. Schwabe (Eds.), *Engineering the Web in the Big Data Era: 15th International Conference, ICWE 2015, Rotterdam, The Netherlands, June 23-26, 2015, Proceedings* (pp. 374-391). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19890-3_24

Lewis, J. R. (1982). Testing small system customer set-up. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 26*, 718-720. https://doi.org/10.1177/154193128202600810

Michailidou, E. (2006). ViCRAM: Visual Complexity Rankings and Accessibility Metrics. *SIGACCESS Access. Comput.*(86), 24-27. http://doi.acm.org/10.1145/1196148.1196154

Michailidou, E., Harper, S., & Bechhofer, S. (2008). *Visual Complexity and Aesthetic Perception of Web Pages*. Paper presented at the Proceedings of the 26th Annual ACM International Conference on Design of Communication, New York, NY, USA. http://doi.acm.org/10.1145/1456536.1456581

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Milic-Frayling, N., & Sommerer, R. (2002). *SmartView: Flexible Viewing of Web Page Contents*. Paper presented at the Poster Proceedings of the Eleventh International World Wide Web Conference. http://www2002.org/CDROM/poster/172/index.html

Needleman, S. B., & Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology, 48*(3), 443-453

Nielsen, J., & Landauer, T. K. (1993). *A Mathematical Model of the Finding of Usability Problems*. Paper presented at the Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, New York, NY, USA. http://doi.acm.org/10.1145/169059.169166

Owens, J. W., Chaparro, B. S., & Palmer, E. M. (2011). Text Advertising Blindness: The New Banner Blindness? *J. Usability Studies, 6*(3), 12:172-112:197. http://dl.acm.org/citation.cfm?id=2007456.2007460

Pernice, K., & Nielsen, J. (2009). *How to Conduct Eyetracking Studies*. Retrieved from https://media.nngroup.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf

Sanoja, A., & Gançarski, S. (2013). Block-o-Matic: a Web Page Segmentation Tool and its Evaluation. In: 29`eme journ´ees "Base de donn´ees avanc´ees", BDA'13. https://hal.archives-ouvertes.fr/hal-00881693

Sanoja, A., & Gançarski, S. (2014). *Block-o-Matic: A web page segmentation framework*. Paper presented at the 2014 International Conference on Multimedia Computing and Systems (ICMCS). http://doi.acm.org/10.1109/ICMCS.2014.6911249

Schmettow, M. (2012). Sample Size in Usability Studies. *Commun. ACM, 55*(4), 64-70. http://doi.acm.org/10.1145/2133806.2133824

Spool, J., & Schroeder, W. (2001). *Testing Web Sites: Five Users is Nowhere Near Enough*. Paper presented at the CHI '01 Extended Abstracts on Human Factors in Computing Systems, New York, NY, USA. http://doi.acm.org/10.1145/634067.634236

Spool, J. M., Scanlon, T., Snyder, C., Carolyn, W., & DeAngelo, T. (1999). *Web Site Usability: A Designer's Guide*: Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. 1-55860-569-X

Takeuchi, H., & Habuchi, Y. (2007). *A Quantitative Method for Analyzing Scan Path Data Obtained by Eye Tracker*. Paper presented at the 2007 IEEE Symposium on Computational Intelligence and Data Mining. http://doi.acm.org/10.1109/CIDM.2007.368885

Tobii Technology AB. (2010). Tobii Studio™ 2.X User Manual (Sep. 2010): Tobii Technology AB.

Toh, H. J., & Hong, J. L. (2014). Webpage Segmentation Using Ontology and Word Matching. In C. K. Loo, K. S. Yap, K. W. Wong, A. Teoh, & K. Huang (Eds.), *Neural Information Processing: 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part II* (pp. 454-461). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12640-1_55

Underwood, G., Humphrey, K., & Foulsham, T. (2008). Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience. In *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work* (pp. 125-144). Berlin, Heidelberg: Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-89350-9_10

Wei, T., Lu, Y., Li, X., & Liu, J. (2015). *Web page segmentation based on the hough transform and vision cues*. Paper presented at the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA). https://doi.org/10.1109/APSIPA.2015.7415394

Yesilada, Y. (2011). *Web Page Segmentation: A Review*. Retrieved from http://emine.ncc.metu.edu.tr/deliverables/emine_D0.pdf

Yesilada, Y., Harper, S., & Eraslan, S. (2013). *Experiential Transcoding: An EyeTracking Approach*. Paper presented at the Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility, New York, NY, USA. http://doi.acm.org/10.1145/2461121.2461134

Zeleny, J., Burget, R., & Zendulka, J. (2017). Box clustering segmentation: A new method for vision-based web page preprocessing. *Information Processing & Management, 53*(3), 735 - 750. http://www.sciencedirect.com/science/article/pii/S0306457316301169

Journal of Eye Movement Research
10(4):6

Eraslan, S., Yesilada, Y., & Harper, S. (2017)
Less Users More Confidence: How AOIs Don't Affect Scanpath Trend Analysis

Appendix

The Web Pages Used in Our Eye Tracking Study

Figure 13. The Apple web page

Figure 14. The Babylon web page
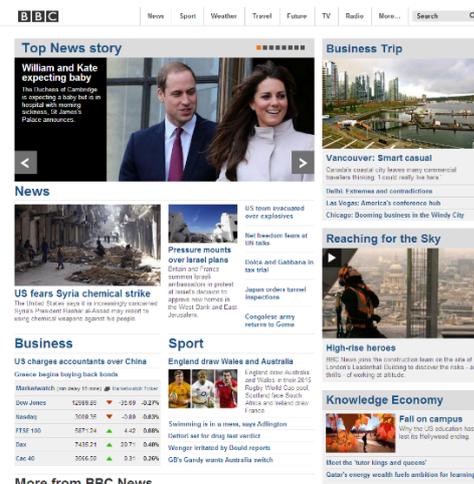

Figure 15. The AVG web page

Figure 16. The Yahoo web page

Figure 17. The Godaddy web page

Figure 18. The BBC web page