Searching on the Go: The Effects of Fragmented Attention on Mobile Web Search Tasks

Morgan Harvey
CIS Department, Northumbria University
Newcastle upon Tyne, United Kingdom
morgan.harvey@northumbria.ac.uk

Matthew Pointon
CIS Department, Northumbria University
Newcastle upon Tyne, United Kingdom
m.pointon@northumbria.ac.uk

ABSTRACT

Smart phones and tablets are rapidly becoming our main method of accessing information and are frequently used to perform on-the-go search tasks. Mobile devices are commonly used in situations where attention must be divided, such as when walking down a street. Research suggests that this increases cognitive load and, therefore, may have an impact on performance. In this work we conducted a laboratory experiment with both device types in which we simulated everyday, common mobile situations that may cause fragmented attention, impact search performance and affect user perception.

Our results showed that the fragmented attention induced by the simulated conditions significantly affected both participants' objective and perceived search performance, as well as how hurried they felt and how engaged they were in the tasks. Furthermore, the type of device used also impacted how users felt about the search tasks, how well they performed and the mount of time they spent engaged in the tasks. These novel insights provide useful information to inform the design of future interfaces for mobile search and give us a greater understanding of how context and device size affect search behaviour and user experience.

CCS CONCEPTS

•**Information systems →Information retrieval;** *Users and interactive retrieval;* •**Human-centered computing →Empirical studies in ubiquitous and mobile computing;** *User studies;*

KEYWORDS

mobile search; fragmented attention; search experience; cognition; user study; experimentation

ACM Reference format:
Morgan Harvey and Matthew Pointon. 2017. Searching on the Go: The Effects of Fragmented Attention on Mobile Web Search Tasks. In *Proceedings of ACM SIGIR conference, Tokyo, Japan, August 2017 (SIGIR 2017),* 10 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recent years have seen rapid growth in the sale and use of various mobile computing devices, giving people the ability to access the Internet away from the confines of a desk, and in many different environmental contexts. Over two-thirds of Americans own a smart phone and almost half own a, somewhat larger, tablet device. At the same time, the sales of desktop and laptop computers have begun to stagnate and even to fall [1]. Almost all smart phone owners (97%) use their devices to access the Internet, many of whom search for information, and to complete fairly complex retrieval tasks: 62% have used them to look up information about a health condition; 57% to do a search for real estate and 40% to look up government services [35].

People use mobile devices to search the web in a variety of different contexts - on public transport, while walking from place to place [17, 23, 32] or in social contexts, where the presence of others can cause distraction [8]. Interaction with such devices is achieved via touch screens upon which small "soft buttons" are drawn for users to select items and input text. Although these buttons may be easy to accurately press in an *ideal environment*, e.g. when seated, such small and non-tactile targets can be significantly more difficult to interact with in other situations [4]. While the ability to perform such tasks "on the go" can be of real benefit, hazards and other changes in the surroundings do necessitate the user's brain switching attention between the ambient environment and the device [11].

These distractions can preoccupy users [30], reducing their effectiveness in interacting with the UI [4, 23] and may even affect user perceptions of the environment and tasks [9]. The result is a larger number of misspelled queries and an attempt by users to shorten queries when searching [32, 33]. In fact, concentration on a mobile task while walking even has an effect on how we walk; to compensate the brain subtly (and subconsciously) alters stance and gait [34]. As such, using a mobile device whilst walking requires both cognitive and motor abilities and so users must divide their attention between the two tasks [21], meaning either an increase in cognitive load, a decrease in pace, a decrease in task performance or a combination of these [22]. The level of difficulty experienced may additionally be influenced by the device size and type and the amount of encumbrance it itself causes [5, 12].

Despite the popularity of mobile devices, their ubiquity in everyday life and the ability they give us to engage in complex search tasks, little is known about how using them on the go impacts upon search behaviour and search performance and whether or not device type and size is an important factor. With this in mind, we investigate whether the small behaviour changes identified in the literature for simple tasks (such as tapping on a highlighted button) result in significant behavioural changes, different perceptions of the task, and different task performance for relatively complex web search problems on both smart phone and tablet devices. Does the change in context impact on user *behaviour*, is this something that

SIGIR 2017, Tokyo, Japan

DOI: 10.1145/nnnnnnn.nnnnnnn

users themselves are aware of and does the type of device used matter? To ensure repeatability, we conducted our study in a lab using simulated contexts - walking on a treadmill, navigating an obstacle course and sitting still at a desk.

Our main research questions, therefore, are:

- Do common mobile situations that cause fragmented attention have an impact on:
  - *RQ1* Users' perceptions of the task and their own performance?
  - *RQ2* Objective measures of users' task performance and behaviour?
- *RQ3* What impact does the device type have on user performance and perception thereof?

The remainder of the paper is structured as follows: In section 2 we consider related work on the topics of mobile device use, fragmented attention and user distraction; section 3 describes the user studies we performed to investigate searching on the go; sections 4,5 and 6 describe the results of the user studies in detail; section 7 discusses how the results relate to the existing literature and suggests reasons and intuition behind them; and section 8 concludes the paper with suggestions for potential future work.

2 RELATED WORK

Improvements in mobile technologies in recent years have led to a dramatic change in how and when people access and use information, and has "a profound impact on how users address their daily information needs" [7]. Research shows that as the power of these devices - as well as the amount of screen space they afford - increases, the complexity of tasks people use them for also increases, with mobile search sessions becoming longer and less homogeneous [19]. Many people now use their smart devices in different contexts to find information, keep up to date with news or to alleviate boredom [35] and frequently use them whilst walking or on public transport. This relatively novel situation of interacting with a computing device when non-stationary can be distracting as attention must be shared (or "fragmented") between operating the device and maintaining motility, typically necessitating a change in posture, stance and gait [34].

A large body of work has investigated user contexts and how fragmented attention affects user input on mobile devices. Early work designed and evaluated forms of human computer interaction in fixed, non-fragmented contexts of use, in a single domain such as a lab [16]. As mobile research evolved, studies began to investigate situations in which attention is diverted from the interface. Oulasvirta et al. found that when following a pre-defined, but otherwise uncontrolled, route through a city users experienced significant impairment when compared with a "non-social laboratory condition" [30]. In a more controlled set of experiments, Lin et al. [23] demonstrated that error rates of stylus input significantly increased as the amount of distraction, and thus degree of attention fragmentation, increased. Similar effects were later demonstrated for touch-based input, with error rates increasing in line with walking speed [28].

Early investigations of reading comprehension and word search when walking [3] showed that contextual variations can have large effects on user behaviour, impairs performance and increases task workload. Mizobuchi et al. looked into mobile text entry and found additional workload effects when walking and identified walking speed as a secondary measure of mental workload [24]. They concluded that texting whilst walking results in either a reduction in input speed (but not accuracy) or a reduction in walking pace. Large-scale analysis of mobile search logs [18] has shown that the increase in time required for mobile searches deters some types of search behaviour, such as exploratory search, and causes search sessions to be considerably shorter than in desktop search. These lines of investigation concluded that times increased significantly when walking compared to a sitting condition, search behaviour altered whilst mobile and walking speed when texting reduces by a fixed amount independent of the level of input difficulty, which varied between participants. These types of investigative conditions create situational impairments which fragment a users' attention, exerting a range of effects on performance and creating compelling opportunities for research [20].

Interaction with such devices is commonly achieved via touch screens upon which relatively small "soft buttons" are drawn for users to select items and input text. The examination of soft buttons, hardware buttons, and surface gestures under conditions of medium and high distraction found that marking menus (i.e. directional gestures) activated along a smartphone's bevel provided the fastest response time [4, 26]. While these buttons may be easy to accurately press in an ideal environment, such as when seated, such small and non-tactile targets may be much more difficult to interact with in other distracting situations [4]. Other investigations assessed the effects of walking on performance with soft buttons, attempting to quantify the negative effects on use due to walking and exploring design changes that may improve a user's experience with a mobile device [20].

Screen real-estate on a mobile device also creates interaction difficulties as a user moves, combined with increasing complexity of mobile task, resulting in considerable obstacles [5, 6, 13]. The limited input modalities afforded by mobile devices have a negative effect on usability [13], a problem compounded by screen size and the device's reduced ability to present information and navigational cues [5, 6]. Small screens can easily become cluttered with information and widgets (buttons, menus, windows, etc.) and this presents a difficult challenge for interface designers [5]. Use of larger devices, such as tablets, which have correspondingly larger screens, may mitigate some of these issues and result in notably different modalities of use [25].

Research shows that smart phones and tablets are often used for different tasks [25, 31] and an analysis of query logs [36] suggests that querying behaviour differs between tablet and smart phone users. Furthermore, there may be a negative correlation between screen size and perceived task difficulty and experienced workload [12], although it has not been investigated when comparing smart phones and tablets and it is unknown what effect situational context has, if any. In general, little is known about the impact differences between the devices has on user behaviour, perceptions and performance on retrieval tasks and under varying mobile conditions.

Delays and time pressures, which may be induced by increased levels of distraction and input error rate, also have a significant impact on search behaviour and objective performance. A study by

Crescenzi et al. [10] compared two groups of users on a number of search tasks: one group was given a per–task time limit of 5 minutes, while the other was given no limit. The results showed that users faced with time pressures experience increased (perceived) task difficulty and less satisfaction with their performance and felt an increased need to work fast and engage in more metacognitive monitoring. Earlier work [9] by the same authors showed that time pressure leads to more queries being issued, fewer documents being viewed and less focus on examination of documents and SERPs. Recent work [15] has demonstrated that users perceive a similar increase in search task difficulty and reduction in satisfaction of their own performance when put under more distracting experimental conditions. These effects are likely as a result of the increased cost of complex cognitive tasks under such conditions, leading to a modification in behaviour as explained by the search models and studies of Azzopardi et al. [2].

Indeed, distractions during walking, driving, and other real-world interactions can preoccupy users [30], reducing their effectiveness in interacting with the UI [4, 23] and resulting in a larger number of misspelled queries and an attempt by users to shorten queries [32, 33]. Walking whilst using a mobile device requires both cognitive and motor abilities and users must divide their attention between the two tasks [21]. This means either an increase in cognitive load, a decrease in pace, a decrease in task performance or a combination of these [22]. There are many examples of distracted input on smart phones where users must split their attention between the task of navigating their physical environment and navigating information on the smart phone screen [26]. It could even be interpreted that users are performing tasks inside a bubble, flipping back and forth between the information on the screen and the outside world [17]. Given that today's users are more likely to be mobile when they search for information online, a deeper understanding of their interactions and challenges whilst mobile will help understand situational search behaviour and the influences of these fragmentations on search.

3 METHOD

We conducted a laboratory experiment with 24 participants drawn from a large European University (a mixture of academic staff, support staff and post-graduate students), of whom 13 were male. Although participants were randomly assigned to one of the 3 conditions, there was a very equal spread of genders with no fewer than 3 of each gender assigned to all conditions ($\mathbb{X}^2$=0.59, p-value=0.75). Ages ranged from 18 to 60, with 2 modal age ranges of between 25 and 30 and between 31 and 40. Ages were also distributed between the experimental conditions with no significant differences ($\mathbb{X}^2$=5.13, p-value=0.74). 18 of the participants were native English speakers and the rest were completely fluent in the language.

There were two independent variables: the type of *device* (tablet or phone; a Huawei MediaPad M2 8" and Moto X Style respectively, both running Android version 5 with the Google Chrome web browser) and the level of *distraction*. The distraction level was varied by simulating 2 everyday situations experienced by mobile device users: walking quickly on a treadmill and navigating an environment with obstacles, as well as a baseline condition in which the participant was seated without any distractions. Participants were randomly allocated to one of the three conditions, resulting in 8 participants for each. Distraction level was a between-subjects variable, while device type was within-subjects.

Following the procedure of Lin et al. [23], participants on the treadmill were asked to select a comfortable walking pace using the increase and decrease belt speed buttons, which was then increased by 20% to induce a small amount of ambulatory distraction. The resulting speeds ranged between 2.2 MPH (3.5 KPH) and 3.8 MPH (6.1 KPH) with a mean of 2.9 MPH (4.7 KPH) and men choosing to walk, on average, 0.78 MPH faster than women. The obstacle course group was shown how to navigate a pre-defined layout (see Figure 1), were asked to maintain a normal walking pace and were prompted to speed up by the researchers if their pace began to noticeably decrease during the task.

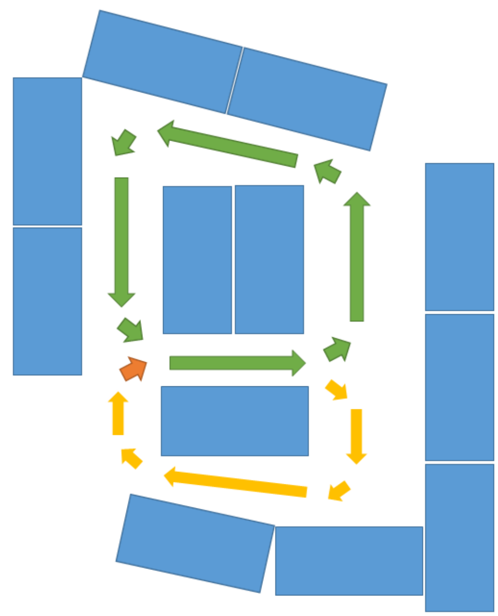

Figure 1: A plan view of the obstacle course layout. Participants began at the orange arrow and followed the course in an anti-clockwise direction.

In order to ensure that we could control the search system and record interaction data we developed a simple mobile search interface named *zing*, shown in Figure 2. The zing interface mimics a standard search engine by showing the titles of 10 links in descending order of relevance together with snippets for each. The interface allowed participants to enter search terms and indicate (via checkboxes) which documents they thought were relevant. It showed the current task (TREC topic) at the bottom of the screen and allowed participants to progress to the next topic at any time. The interface also prompted users to fill in pre- and post-topic questionnaires to survey their perceptions about the task and their self-assessed post-task performance, satisfaction, perceived time pressure and focus/involvement on the task. Half of the participants completed their first 2 topics on a phone, moving on to the tablet for their final 2 topics, while the other half began with the tablet.

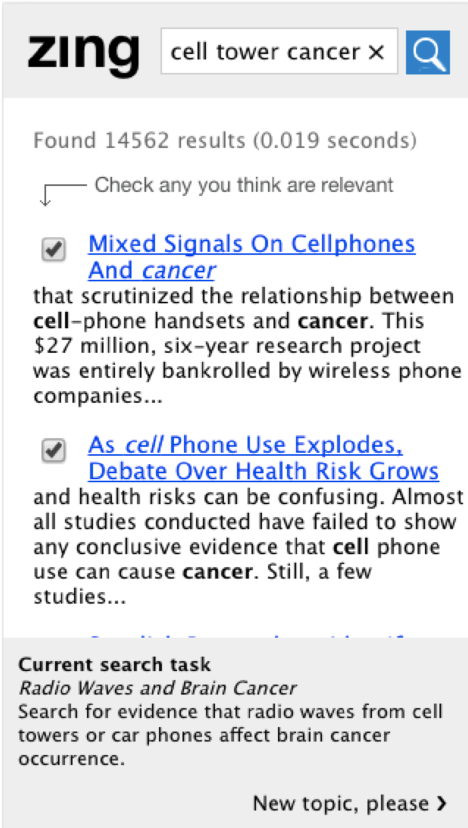

Figure 2: zing search interface on an Apple iPhone 5. Checkboxes used to indicate relevance.

#	Title	AP	Pre	Post
362	Human smuggling	0.29	2.83	2.75
367	Modern Piracy	0.26	2.79	2.25
638	Wrongful convictions	0.23	2.83	3
404	Ireland peace talks	0.28	3.25	2.79

Table 1: TREC topics used.

We used a standard test collection: AQUAINT, and removed duplicate documents in a pre-processing step to provide a better and more familiar user experience. To assess performance we made use of pre-defined TREC topics from the 2005 Robust track [37], of which we chose 4 at random from a subset of those which are neither too difficult nor too easy[1]. Table 1 shows the topics chosen as well as the average precision (AP) of their titles on the AQUAINT collection and the participants' perceptions of each topic's difficulty before (pre) and after (post) completing it.

[1]After the method of Harvey et al. [14], whereby the difficulty of a topic is determined by the average precision of its title over the document collection.

Indexing, searching and snippet generation was provided by Apache SOLR[2]. Each participant was given the same 4 topics (tasks) in a random order with a per-task time limit of 15 minutes and alternated between the two *device* conditions by conducting the first two tasks on one device before switching to the other for the final 2 topics. The starting device for each user was allocated at random to prevent fatigue and/or learning effects from confusing the results. Participants were asked to imagine they wanted to learn more about the subject of each topic for a short report and were requested to select between two and four documents they thought were relevant for each topic and were told they could submit multiple queries per topic, if necessary. Participant actions and behaviour were recorded by means of a GoPro camera worn on the head, a wide-angle view of the obstacle course and by recording and logging interactions with the touchscreen and browser interface (Figure 3).

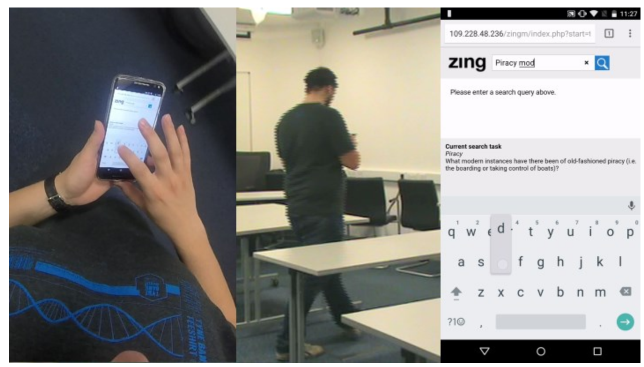

Figure 3: Example of data recorded via the cameras and screen recording software. Note that information from all 3 sources is temporally synced.

4 RESULTS

In the following we use t-tests to compare distributions that are normal (as well as results from Likert scales) and Wilcoxon sign-rank tests in cases of non-normal data (e.g. task duration and number of hits).

4.1 Pre-study questionnaire

Before being told anything about the experiment, participants were asked to fill in a short pre-study questionnaire asking them about their use of mobile devices and search engines as well as how difficult they would expect it to be to search on a phone or a tablet in various contexts.

All but two participants use a mobile device several times a day and all but three use a search engine to find information several times per day and all participants but one said they were either "confident" or "very confident" at using a search engine to find information. 19 use their mobile device at least once per day whilst walking, 9 use it daily on public transport and all but 3 use it to search the web on a daily basis. Participants expected that using both devices whilst walking on a treadmill, navigating an obstacle

[2]http://lucene.apache.org/solr/

course or while sitting in a noisy pub or cafe would be significantly more difficult than when sitting still (see Figure 4). They expected using a mobile phone to be significantly more difficult when navigating an obstacle course compared with when walking on a treadmill (t=2.95, p-value=0.005) and expected, for both devices, that searching in a noisy pub or cafe would be significantly easier than in either of the other two conditions (all tests p-value $\ll 0.01$).

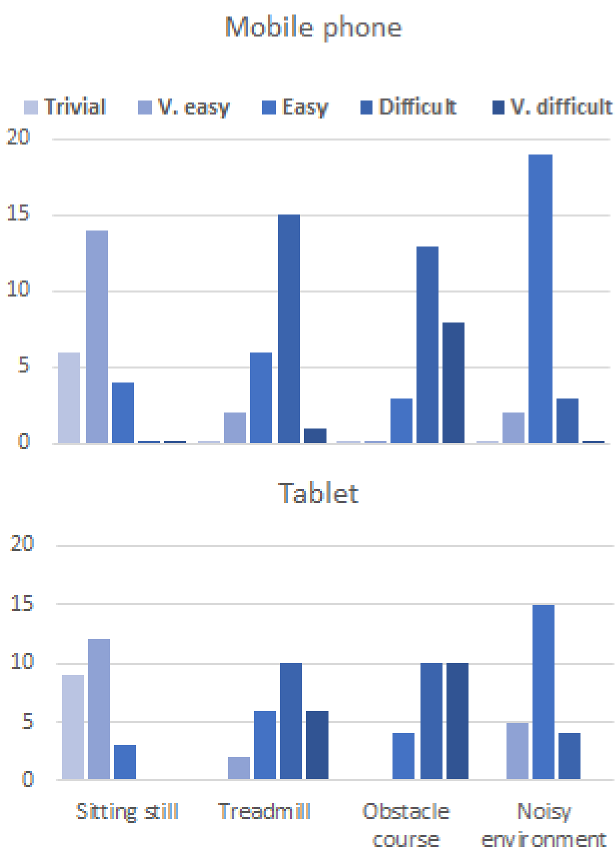

Figure 4: Expected difficulty of searching on mobile phones and tablets under various conditions.

As participant age increased, the expected difficulty of using either a mobile phone or a tablet on a treadmill (R-squared=0.27, p-value=0.005; R-squared= 0.17 p-value=0.028) and when navigating an obstacle course (R-squared=0.34, p-value=$\ll$0.01; R-squared= 0.29, p-value=0.004) increased, however this was not the case for use when sitting still or in a noisy pub or cafe. The more confident people were at using search engines in general, the easier they expected the task to be on the treadmill (R-squared=0.24, p-value=0.015) and the obstacle course (R-squared=0.2, p-value: 0.03) on both devices. However, this relationship only held for the tablet when imagining sitting still (R-squared=0.28, p-value=0.008). There was no significant relationship between search engine confidence and expected difficulty in the noisy pub environment. Surprisingly, the participants' familiarity of using mobile devices when walking or in noisy environments was not predictive of their expected difficulty of searching under the same conditions.

4.2 Pre-task perception

Before each task (TREC topic), the zing interface prompted participants to fill in a short questionnaire about their prior knowledge of the topic, their interest in it and how difficult they expected the task to be (overall difficulty, difficulty in finding relevant documents, and difficulty in knowing when to finish; see Figure 5). To aid them in doing so, the topic title and description were presented at the bottom of the screen. There was little variation in the responses between the topics with most people stating that they had fairly little prior knowledge and were moderately interested in the topics. Responses did indicate an expectation that topic 404 ("Ireland peace talks") would be the most difficult, although the difference was not significant. There were only two instances where a participant was unsure of how to complete the task and in only 14% of cases was a topic deemed to be either very difficult or very easy. As expected, responses to all 3 questions on perceived task difficulty were all significantly correlated with each other.

Figure 5: Results of pre-task questionnaire.

Condition	Sitting	Obstacles	Treadmill
Overall difficulty	2.21	3.06	3.36
Finding rel. docs.	2.43	2.59	3.03
When to finish	2.79	3.06	3.58

Table 2: Mean responses about task difficulty from pre-task questionnaires by condition.

It seems that participants took experimental condition into account when estimating the difficulty of tasks as there were differences in the perceived difficulty of tasks, as shown in Table 2. Those who knew they would be sitting still expected the tasks to be significantly easier than those who were navigating the obstacle course (t=3.95; p-value $\ll$ 0.01) and those who were on the treadmill (t=5.08; p-value $\ll$ 0.01). Those who were sitting still and those on the obstacle course thought finding relevant documents would be equally easy (t=0.7, p-value=0.49), however those on the treadmill expected this to be significantly more difficult (t=2.58, p-value=0.012). The treadmill group thought that knowing when to finish the task (i.e. ascertaining when they'd found enough information) would be significantly more difficult than the baseline group (t=3.15, p-value=0.002). There were no significant differences in perceived task clarity between any of the groups, although those in the baseline group did claim to know more about the topics a priori than those in the other groups (compared to treadmill: t=2.22, p-value=0.031 ; compared to obstacle course: t=2.18, p-value=0.033).

4.3 Post-task perception

#	Question
Q1	I felt hurried or rushed when completing this task
Q2	It was important to complete this task quickly
Q3	Overall, I thought this was a difficult task
Q4	I am satisfied with steps I took to find information
Q5	I forgot my immediate surroundings during the task
Q6	I was so involved that I ignored everything around me
Q7	I was so involved that I lost track of time
Q8	I was absorbed in my search task
Q9	I found enough info. about the search topic
Q10	I am satisfied with the info. I found

Table 3: Selected post-task questions.

Immediately after each task participants filled in a post-task questionnaire, which included items from the *focused attention scale* of O'Brien et al. [29] as well as items from Crescenzi et al. [10] (see Table 3 for selected items). The questions were chosen to ascertain the participants' levels of perceived time pressure, self-assessed performance and involvement in the search task. There were significant differences in terms of perceived difficulty between the 4 topics with 2 topics scoring a median Q3 ("Overall, I thought this was a difficult task") agreement of 2, one at 3 and the most difficult scoring 4. There were, however, no significant differences between the 4 topics for the other questions. Interestingly, women reported feeling significantly less absorbed in the task (Q8; t=2.96; p-value=0.004) than men and felt less like they lost track of time (Q7; t=1.99; p-value=0.049).

As shown in Table 4, the different experimental conditions had a number of different effects on the participants' perceptions. Those on the treadmill felt significantly more rushed than in the other two conditions (Q1) and those sitting still felt significantly less pressure to complete the tasks quickly than the other 2 groups (Q2).

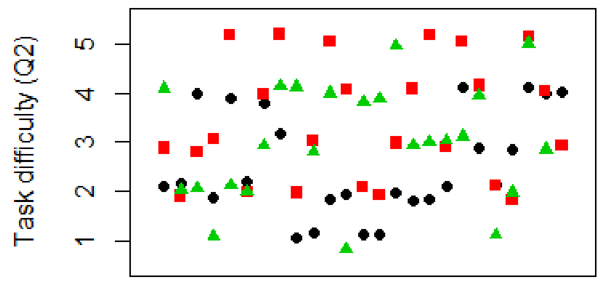

Figure 6: Perceived post-task difficulty by condition. ● = sitting; ▲ = obstacle course; ■ = treadmill

Condition	Sitting	Obstacles	Treadmill
Q1	2.25 †	2.53 †	3.28
Q2	2.14 *†	2.87 †	3.44
Q3	2.43 †	3.0	3.31
Q4	3.86 †	3.47	3.03
Q5	3.64 *	3.16 †	3.42 *
Q6	3.53	3.03 †	3.47 *
Q7	3.39	3.09	3.52
Q8	3.96	3.56	3.81
Q9	3.39 †	3.75 †	2.91
Q10	3.46 †	3.56 †	2.8

Table 4: Mean responses from post-task questionnaires by condition. * = sig. diff. with Obstacles; † = sig. diff. with Treadmill

It appears that those sitting still generally found the tasks easiest (Q3; see Figure 6) - significantly more so than those in the treadmill group - and were more satisfied with the steps they took to find relevant information (Q4). Those sitting and on the treadmill were significantly more likely to forget their immediate surroundings than those on the obstacle course (Q5) and felt more involved in the task (Q6). Although differences were not significant, there was a trend that those on the treadmill felt more involved in the task to the point where they lost track of time (Q7) and those on the obstacle course felt less absorbed in the search tasks (Q8). In terms of being able to find sufficient information to fulfill the task, those in the baseline and obstacle course conditions felt there significantly more able to find enough information (Q9) and were significantly more satisfied with what they found than those on the treadmill (Q10).

5 SEARCH PERFORMANCE

In order to objectively evaluate search performance, we rely on three main metrics: the average number of hits (relevant documents) returned per search query; the mean average precision attained; the number of documents bookmarked; the number of documents read; the ratio of relevant documents bookmarked relative to the total number bookmarked (to give an indication of how accurate users were with their bookmark choices); and the same ratio for documents read. Based on the results of linear models, the number of hits, mean average precision and number of documents read are all significant predictors of perceived success (Q9 and Q10 in the post-task questionnaire). We also consider a number of other proxies of overall search and task performance as well as metrics such as query length and search duration.

5.1 Performance by experimental condition

Condition	Sitting	Obstacles	Treadmill
# of queries/user	13	12	14
Hits/query	3.71 ∗†	2	1.75
MAP	0.104 ∗†	0.085	0.083
Bookmarks/query	1.32 †	1.74 †	1.03
(Ratio relevant)	0.55	0.47	0.49
Docs read/query	1.58 †	1.19	1.0
(Ratio relevant)	0.43	0.41	0.44
# of query terms	3.61 ∗	3.17	3.38
Query duration	39.5s ∗†	30.5s	35s

Table 5: Objective performance measures by condition. ∗ = sig. diff. with Obstacles; † = sig. diff. with Treadmill

Table 5 shows how the objective performance measures varied by experimental condition. Most notably, the average number of hits per query achieved by the baseline users is significantly greater than those by either the treadmill (p-value=0.029) or obstacle course (p-value=0.023) groups, even though all groups submitted very similar numbers of queries (see Figure 7). This is also true for mean average precision. This suggests that those sitting were able to generate more accurate and precise queries than those in the other two groups. This may be because the queries they submitted were longer and more detailed (significantly longer than the obstacle course group: p-value=0.002) and because they spent significantly more time per query than the others - over 5 seconds longer on average per query (compared to treadmill: p-value=0.023; compared to obstacle course: p-value=0.005).

Those sitting and those on the obstacle course bookmarked significantly more documents than the treadmill group (p-values= 0.01 and 0.001 resp.). The participants on the obstacle course bookmarked the most often, however, as they bookmarked a larger number of non-relevant documents, they had the lowest ratio of relevant bookmarks. The baseline group read the largest number of documents on average, perhaps partially explaining their increased query durations, and read significantly more than those on the treadmill (W=7371, p-value= 0.015). This may be because sitting at a desk is a more comfortable environment for in-depth tasks such as reading, which requires concentration and may be disrupted by movements of the screen or eyes.

6 IMPACT OF DEVICE USED

To determine what impact device type has on search, half of the search tasks were completed on a smart phone and the other half were completed on a larger tablet device. As shown in Table 6, although the objective performance measures recorded for the different devices were almost identical (i.e. no significant differences), there was substantial variation in the participants' perceptions of

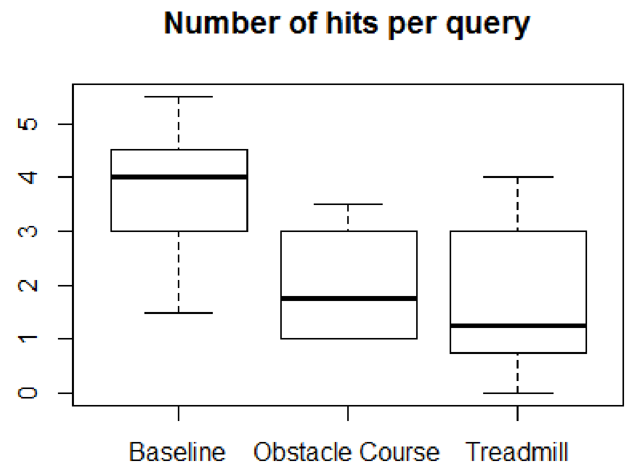

Figure 7: Number of hits (relevant documents) returned per query.

searching on each device. In general, people found the smart phone to be much less useful for the tasks set than the tablet: They felt significantly more hurried and rushed when using the phone (t=2.25; p-value=0.025) and found the tasks to be significantly more difficult (t=2.7; p-value=0.007). Although users felt equally satisfied on both devices about the steps they themselves had taken to find the necessary information (t=-0.45; p-value=0.65), when using the smart phone they were significantly less satisfied with the information they found (t=-3.14; p-value $\ll$ 0.01), suggesting that they placed the blame on the device and not on their own search behaviour.

Device type	Smart phone	Tablet
Hits/query	2.8	2.77
Bookmarks/query	1.48	1.2
# of query terms	3.39	3.4
Query duration	48.6	49.2
Q1 felt hurried/rushed	2.98	2.69
Q3 difficult task	3.49	3.12
Q4 satisfied with step taken	3.18	3.25
Q10 satisfied with info. found	2.77	3.22

Table 6: Objective and subjective performance measures by device type.

It seems the experimental condition had an impact on how users perceived differences between the devices (Table 7). Users in the baseline condition (sitting at a desk) actually performed better - in terms of number of hits - on the tablet than on the phone, albeit not significantly (W=977, p-value=0.121). This trend was, however, reversed under the other two experimental conditions with those on the phone seemingly performing better than those on the tablet. This was also reflected in the users' perception of flow/involvement in the task: Those sitting felt significantly less aware of their surroundings when using the tablet (Q5; t = 2.2, p-value=0.03) than the phone, while those in the other conditions had the opposite experience (Q5; t=-2.11, p-value=0.036); and those in the non-baseline conditions felt less aware of time passing when using the phone than the tablet (Q7; t=3.53, p-value=0.001). It's also notable that the baseline group spent longer on the tasks (*query duration*) when using the tablet than the phone, but the other groups actually spent longer when using the phone.

	Baseline		**Other cond.**	
	P	***T***	***P***	***T***
Hits/query	2.67	3.64	2.84	2.37
Query duration	59.1	64.1	44.2	42.2
Q5 forgot surroundings	3.17	3.7	3.53	3.18
Q7 lost track of time	2.98	3.42	3.73	3.1

Table 7: Performance and perception by condition and device type (P=smart phone, T=tablet).

7 DISCUSSION

This research set out three research questions aimed at exploring mobile searching and the effects of fragmented attention in common situations. The following discussion will consider each of the research questions in turn.

Do common mobile situations impact on users' perceptions of the task and their own performance? (RQ1)

Our results demonstrate that the different conditions had a number of fairly profound effects on user perceptions, both before and after completing the tasks. The pre-study questionnaire showed that participants expected using both devices whilst walking on a treadmill would be more difficult than sitting still and navigating an obstacle course. This is something that tallies with past research, which shows that situational impairments do exert a range of effects on performance, adding levels of difficulty as interaction with the device takes place [20]. The treadmill lessened their feeling of control, or lack of it, which reduced their perceived effectiveness as they interact with the UI [4, 23]. The older a participant was, the greater the expected difficulty of using a tablet on a treadmill, but this was not the case for phones or when sitting, perhaps because younger people are more familiar with such devices and may have more experience using them in mobile situations [1].

Post-task perception showed that different experimental conditions had a number of different effects on the participants' perceptions. Those on the treadmill felt significantly more rushed than in the other two conditions. Oulasvita et al. [30] pointed to the effect of a situation on the duration of continuous attention, finding that participants in their laboratory experiments were more focused on the tasks compared with participants on a busy street. In this study, those sitting and on the treadmill were significantly more likely to forget their immediate surroundings than those on the obstacle course and more involved in the task. This may be because there is an increased need to attend to the surrounding environment when walking, but with the treadmill this is not the case as the situation does not change [23].

Participants seemed to take the experimental conditions into account when estimating task difficulty, recording significant differences in perceived task difficulty. With the frequency of mobile use continuously on the increase, participants were likely to be aware of these potential challenges as they interacted. They expected these difficulties to increase their cognitive workload and the changes in mobility (i.e. walking) to influence not only their walking speed but mental workload during the tasks [24]. Those who knew they would be sitting still expected the tasks to be easier than the other conditions while those who were sitting still and those on the obstacle course thought finding relevant documents would be equally easy.

It is interesting that people expected the treadmill to be most difficult, despite the fact that it should require more cognitive effort to avoid the obstacles. This may be because these participants have control over the pace at which they are walking, while those on the treadmill are kept at a constant speed by the mechanism. Those on the obstacle course have the possibility to slow down while conducting demanding tasks, such as assessing document relevance, thereby reducing their overall cognitive load [21]. This may explain why Mizobuchi et al. [24] observed no reduction in input accuracy when walking and texting - the participants simply reduced their walking speed to prioritise text input.

Participants on the obstacle course felt less absorbed in the search tasks. This could be due to the fact that walking while using a smart phone requires both cognitive and motor abilities and appropriate division of attention to each [20]. The level of absorption in the search tasks is less due to the participant needing to be aware of their surroundings. The participants are walking and using the device, in doing so they take longer to complete a set route and, therefore, walk more slowly. There are two repercussions to this, they will slow down on the obstacle route (because they have control) and experience increased cognitive load on the treadmill (not being able to adjust their speed) [22].

Do common mobile situations impact on objective measures of users' task performance and behaviour? (RQ2)

Although the effects on objective performance were perhaps not quite as numerous or great as they were on perception, the different conditions did impact search behaviour and, consequently, performance. The most profound difference was found in the quality, in terms of number of hits and MAP, of the queries submitted - those sitting were able to generate significantly more accurate and precise queries than those in the other two groups. Perhaps this is because sitting evokes an environment more like desktop search, where users feel that they have more time to think carefully about the queries they enter [18]. This was also evidenced by the sitting group's queries being significantly longer (i.e. being comprised of more terms) and is in line with the studies of Kamvar et al. [18] and Schaller et al. [32, 33] and also corresponds with the results from the post-task questionnaire, which showed that the users on the treadmill and on the obstacle course felt more hurried and rushed and were more aware of time pressures.

Additionally, it seems the effects of time pressure on search behaviour highlighted in the studies of Crescenzi et al. [9, 10] are also relevant in this context, even though in the case of our study time pressures were perceived rather than enforced. Interestingly, though, we did not observe the same increase in querying frequency. [9]. A possible way to mitigate these issues might be to detect when users are walking (by using the device's motion sensors and gyroscopes) and to adapt the interface to offer more querying support and to present more concise snippets in such situations.

Participants on the treadmill bookmarked significantly fewer documents than the other two groups. A situation which is again likely because they felt more rushed, meaning they were less likely to explore the search results and to assess potentially relevant documents for relevance [9], tasks that will likely incur a higher "cost" [2] when input accuracy [26] and reading comprehension [3] is reduced. Similarly, participants in both of the non-baseline groups spent significantly less time on each SERP and, therefore, assessed significantly fewer documents for relevance.

What impact does the device type have on user performance and perception thereof? (RQ3)

There was substantial variation in the participants' perceptions of searching on each device, contradicting the objective performance observed on the devices, which were identical. We found that the device used influenced participants' perceptions of the search tasks and that the tablet was, on the whole, preferred, although this was somewhat dependent on experimental condition. People felt more hurried, found the tasks harder and were less satisfied with the information they had found when using the phone. The increased (perceived) difficulty on the phone may be because users have less screen space to work with, making interaction with the various UI controls more difficult, especially when interaction occurs in a distracting environment [4]. Since larger screens appear less cluttered with information, users may have felt less overwhelmed by the amount of information presented on the relatively more spacious tablet screen [5]. These findings are in line with those of Hancock et al. [12], however our results are novel as they demonstrate that this effect holds between smart phone and tablet devices and is in fact more profoundly felt in the context of mobile search.

In contrast to the results of Song et al. [36], we didn't find any difference in query length or query duration between the two devices, although there was notable interaction between the device type and experimental condition. Users in the baseline (seated) group performed better on the tablet than the phone, however, those in the other two groups performed rather better on the phone than the tablet. This may be because phones are more typically used as a handheld device at arm's length, while the larger, heavier tablets are more often used when propped up on a table or cradled in one arm [31] and rarely used out of the home [25]. This is also evidenced by the difference in perceived immersion/flow in the task - when seated, using the tablet resulted in a greater feeling of immersion than the phone, while this was reserved for the other two conditions. The extra heft of the tablet when walking may make the device too conspicuous, serving to pull the user out of flow, while the much lighter, less cumbersome phone does not prevent the users from becoming immersed.

The variation in the amount of time spent on tasks (baseline users spend longer on the tablet than the phone, with the situation reversed for the other conditions) is interesting and perhaps speaks to the difference in weight (and therefore experienced encumbrance) between the two devices. Increased encumbrance has been shown to result in reduced input accuracy and increased mental load [27] and may lead to users more rapidly becoming fatigued, which would explain their propensity to give up the tasks earlier on the tablet when walking. When choosing between devices for a given task, it may therefore be useful to consider whether or not the user is likely to be moving or seated.

8 CONCLUSIONS

The main aim of this study was to investigate how different mobile situational contexts and different mobile devices (i.e. phones and tablets) affect user performance and experience when performing web search tasks. We conducted a laboratory experiment with 24 participants in which three different conditions were simulated: sitting at a table (the baseline), walking on a treadmill and navigating an obstacle course. Analysis of subjective measures, derived from pre- and post-task questionnaires, as well as objective performance metrics showed that both the context and device variables had a number of effects on performance, both perceived and measured, as well as participants' feelings of immersion, satisfaction and urgency.

Our results provide useful insights to inform the design of future interfaces for mobile search and give us a greater understanding of how context and device size affect search behaviour and user experience. It is clear that some contexts have negative effects on user search experience and that this is additionally affected by device type. When seated, tablets are preferable for complex search tasks, however this is reversed in instances where the advantage of the device's extra screen space is offset by its additional weight (and therefore, the extra encumbrance experience by the user).

These insights suggest the need for more care to be taken when designing mobile search interfaces by considering the context in which the system will be used, as well as the type of device. Interfaces could be developed that adapt when a walking-like motion is detected to aid the user in generating queries and to present information in a terser, more focused manner to reduce mental load and simplify the information space. This work also has potential repercussions for IR and HCI researchers: When designing and evaluating mobile search systems, it is clear that whether the user is in motion and the combination of device size and weight and situational context have significant effects on perception. It is also clear from this work that a treadmill may not always be appropriate for simulating mobile search as in reality users adjust their walking speed to prioritise interaction with the device, something which is not possible under this condition. Therefore, practitioners should be aware of these factors to ensure that these insights are incorporated into study design and taken into account when assessing user performance so that results are in fact demonstrative of effects induced by the experimental conditions and not other unmeasured variables.

8.1 Future work

As future research in this area we plan to expand on this work by looking into user search behaviour in more detail using the additional qualitative sources of information we captured during the study. As noted earlier, we have recorded GoPro footage of each participant as well as screen recordings of their interactions which we plan to evaluate to identify patterns and behaviours unique to each experimental condition. Using the data from the GoPro we will be able to evaluate the participants' spatial awareness (especially on the predefined route) and their "attention-switches" away from the device in different situations. Using the 3 everyday situations we will be able to assess the levels of immersion with each task and compare the GoPro data to the pre-task perceptions - does their initial thinking match reality and can we confirm our suspicions that the tablet's weight and bulk is the main cause of the differences observed in this research? We intend to develop search interfaces that adapt to the user's situation (i.e. walking or not) and the device type and to investigate whether these changes can in fact aid users in fragmented contexts to query as well as those who are seated. We would also like to simulate other situations that induce attention fragmentation, such as a busy restaurant or bar, and determine whether or not this causes similar changes in user behaviour and performance.

REFERENCES

[1] Monica Anderson, *Technology device ownership, 2015*, Pew Research Center, 2015.

[2] Leif Azzopardi, Diane Kelly, and Kathy Brennan, *How query cost affects search behavior*, Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, ACM, 2013, pp. 23–32.

[3] Leon Barnard, Ji Soo Yi, Julie A Jacko, and Andrew Sears, *Capturing the effects of context on human performance in mobile computing systems*, Personal and Ubiquitous Computing **11** (2007), no. 2, 81–96.

[4] Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley, *Experimental analysis of touch-screen gesture designs in mobile environments*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2011, pp. 403–412.

[5] Stephen Brewster, *Overcoming the lack of screen space on mobile computers*, Personal and Ubiquitous Computing **6** (2002), no. 3, 188–205.

[6] Minhee Chae and Jinwoo Kim, *Do size and structure matter to mobile users? an empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones*, Behaviour & information technology **23** (2004), no. 3, 165–181.

[7] Karen Church, Mauro Cherubini, and Nuria Oliver, *A large-scale study of daily information needs captured in situ*, ACM Transactions on Computer-Human Interaction (TOCHI) **21** (2014), no. 2, 10.

[8] Karen Church and Nuria Oliver, *Understanding mobile web and mobile search use in today's dynamic mobile landscape*, Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, ACM, 2011, pp. 67–76.

[9] Anita Crescenzi, Diane Kelly, and Leif Azzopardi, *Time pressure and system delays in information search*, Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 767–770.

[10] ———, *Impacts of time constraints and system delays on user experience*, ACM CHI, ACM, 2016, pp. 141–150.

[11] Mark Dunlop and Stephen Brewster, *The challenge of mobile devices for human computer interaction*, Personal and ubiquitous computing **6** (2002), no. 4, 235–236.

[12] PA Hancock, BD Sawyer, and S Stafford, *The effects of display size on performance*, Ergonomics **58** (2015), no. 3, 337–354.

[13] Rachel Harrison, Derek Flood, and David Duce, *Usability of mobile applications: literature review and rationale for a new usability model*, Journal of Interaction Science **1** (2013), no. 1, 1.

[14] Morgan Harvey, Claudia Hauff, and David Elsweiler, *Learning by example: training users with high-quality query suggestions*, Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 133–142.

[15] Morgan Harvey and Matthew Pointon, *Perceptions of the effect of fragmented attention on mobile web search tasks*, ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR), 2017.

[16] Peter Johnson, *Usability and mobility; interactions on the move*, Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices, 1998.

[17] Anne Kaikkonen, *Full or tailored mobile web-where and how do people browse on their mobiles?*, Mobility, ACM, 2008, p. 28.

[18] Maryam Kamvar and Shumeet Baluja, *A large scale study of wireless search behavior: Google mobile search*, Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM, 2006, pp. 701–709.

[19] ———, *Deciphering trends in mobile search*, IEEE Computer **40** (2007), no. 8, 58–62.

[20] Shaun K Kane, Jacob O Wobbrock, and Ian E Smith, *Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces*, Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, ACM, 2008, pp. 109–118.

[21] Eric M Lamberg and Lisa M Muratori, *Cell phones change the way we walk*, Gait & posture **35** (2012), no. 4, 688–690.

[22] Sammy Licence, Robynne Smith, Miranda P McGuigan, and Conrad P Earnest, *Gait pattern alterations during walking, texting and walking and texting during cognitively distractive tasks while negotiating common pedestrian obstacles*, PLoS one **10** (2015), no. 7, e0133281.

[23] Min Lin, Rich Goldman, Kathleen J Price, Andrew Sears, and Julie Jacko, *How do people tap when walking? an empirical investigation of nomadic data entry*, International Journal of human-computer studies **65** (2007), no. 9, 759–769.

[24] Sachi Mizobuchi, Mark Chignell, and David Newton, *Mobile text entry: relationship between walking speed and text input task difficulty*, Proceedings of the 7th international conference on Human computer interaction with mobile devices & services, ACM, 2005, pp. 122–128.

[25] Hendrik Müller, Jennifer L. Gove, John S. Webb, and Aaron Cheang, *Understanding and comparing smartphone and tablet use: Insights from a large-scale diary study*, Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (New York, NY, USA), OzCHI '15, ACM, 2015, pp. 427–436.

[26] Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank, *Tap, swipe, or move: attentional demands for distracted smartphone input*, Proceedings of the International Working Conference on Advanced Visual Interfaces, ACM, 2012, pp. 173–180.

[27] Alexander Ng, John Williamson, and Stephen Brewster, *The effects of encumbrance and mobility on touch-based gesture interactions for mobile phones*, Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, 2015, pp. 536–546.

[28] Hugo Nicolau and Joaquim Jorge, *Touch typing using thumbs: Understanding the effect of mobility and hand posture*, ACM CHI (New York, NY, USA), CHI '12, ACM, 2012, pp. 2683–2686.

[29] Heather L O'Brien and Elaine G Toms, *The development and evaluation of a survey to measure user engagement*, Journal of the American Society for Information Science and Technology **61** (2010), no. 1, 50–69.

[30] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti, *Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci*, ACM CHI, ACM, 2005, pp. 919–928.

[31] Tommaso Piazza, Morten Fjeld, Gonzalo Ramos, AsimEvren Yantac, and Shengdong Zhao, *Holy smartphones and tablets, batman!: mobile interaction's dynamic duo*, Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction, ACM, 2013, pp. 63–72.

[32] R. Schaller, M. Harvey, and D. Elsweiler, *Out and about on museums night: Investigating mobile search behaviour for leisure events*, Searching4Fun WS at ECIR, 2012.

[33] Richard Schaller, Morgan Harvey, and David Elsweiler, *Entertainment on the go: finding things to do and see while visiting distributed events*, IIiX, ACM, 2012, pp. 90–99.

[34] Kelly M Seymour, Christopher I Higginson, Kurt M DeGoede, Morgan K Bifano, Rachel Orr, and Jill S Higginson, *Cellular telephone dialing influences kinematic and spatiotemporal gait parameters in healthy adults*, Journal of Motor Behavior **48** (2016), no. 6, 535–541.

[35] Aaron Smith et al., *Us smartphone use in 2015*, Pew Research Center **1** (2015).

[36] Yang Song, Hao Ma, Hongning Wang, and Kuansan Wang, *Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance*, Proceedings of the 22nd international conference on World Wide Web, ACM, 2013, pp. 1201–1212.

[37] E.M. Voorhees, *The trec 2005 robust track*, SIGIR Forum **40** (2006), no. 1, 41–48.