Article:

Harvey, M. and Pointon, M. (2019) Understanding in-context interaction: An investigation into on-the-go mobile search. Information Processing and Management, 56 (6). 102089. ISSN 0306-4573

https://doi.org/10.1016/j.ipm.2019.102089

Understanding in-Context Interaction: an Investigation into On-the-go Mobile Search

Morgan Harvey and Matthew Pointon

Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, UK

Abstract

Recent years have seen a profound change in how most users interact with search engines: the majority of search requests now come from mobile devices, which are used in a number of distracting contexts. This use of mobile devices in various situational contexts away from a desk presents a range of novel challenges for users and, consequently, possibilities for interface improvements. However, there is at present a lack of work that evaluates interaction in such contexts to understand what effects context and mobility have on behaviour and errors and, ultimately, users' search performance.

Through a controlled study, in which we simulate walking conditions on a treadmill and obstacle course, we use a combination of interaction logs and multiple video streams to capture interaction behaviour as participants (n=24) complete simple search tasks. Using a bespoke tagging tool to analyse these recordings, we investigate how situational context and distractions impact user behaviour and performance, contrasting this with users in a baseline, seated condition. Our findings provide insights into the issues these common contexts cause, how users adapt and how such interfaces could be improved.

Keywords: mobile search; distraction; search experience; user study; experimentation

1. Introduction

Recent years have seen a rapid change in the mix of devices people use to search for and interact with information on the Internet. As early as mid-2015 the world's largest search engine announced that mobile devices had overtaken desktop and laptop machines as the most common source of search queries [15]. Such devices give people the ability to access search engines and the wider Internet away from the confines of a desk and in many different environmental contexts: on public transport, while walking from place to place [28, 34, 46] or in social contexts, where the presence of others can cause distraction [9]. Put simply, we use mobile devices in situations where we cannot use desktops and laptops and these everyday situations (e.g. walking) can often present distractions, dividing our attention as we interact [40]. Research suggests that these situations increase cognitive load and, therefore, may have an impact on performance [23], increasing the possibility of interaction errors as we hold and alter our grip to complete common tasks.

The result is a larger number of misspelled queries and an attempt by users to shorten queries when searching [45, 46]. 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 [47]. 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 [32], meaning either an increase in cognitive load, a decrease in pace, a decrease in task performance or a combination of these [33]. The level of difficulty experienced may additionally be influenced by the device size and type and the amount of encumbrance it itself causes [21, 6].

Since interaction with mobile devices is typically achieved by means of a touchscreen and such devices are used whilst held in one's hand, it is important to consider how the device is gripped [18, 17]. Research shows that people employ various different grips and that, depending on factors such as device size, they have preference of certain grips over others [18]. Despite the fact that mobile hand-held devices are often used in distracting contexts and whilst on the move and for complex tasks, we do not yet know what impact different situations have on grip preference and effectiveness, how users transition between grips as a task evolves and how this all relates to interaction errors. While previous work has investigated behaviour and error rates when completing single "atomic" tasks (e.g. tapping a button or typing a set word or phrase), we know little about user interaction with devices for more holistic tasks, such as web search, and the effects that commonly-encountered mobile situations have on these interactions.

In this paper we investigate user search performance and behaviour, interaction errors and the use of grips in common everyday situations (i.e. walking whilst using a mobile phone). Through a controlled study we asked participants to complete real-world search activities and, using both log data from the devices and multiple video recordings, captured their behaviour and how they interacted with the device in completing the tasks. These recordings were synced, encoded, annotated and analysed using a bespoke tool designed specifically for this research. By manually tagging interactions observed from the three video streams running concurrently, we are able to analyse behaviour and identify which grips are being used, when these are changed and how this impacts search performance. In concert with analysis of log data, these investigations allow us to gain a more nuanced and detailed understanding of device use in context. This research is, to our knowledge, the first to investigate grips in situations where mobile device users are not seated at a desk and interaction errors for complex tasks under varying everyday mobile situational contexts.

Concretely, we aim to:

1. Obtain, from a lab study, holistic data of user behaviour when performing search tasks on a mobile phone in everyday situations.
2. Investigate the effects these mobile situations have on hand movements, grip and shifts in grip.
3. Assess the impact of (i) context and (ii) grip on:
    (a) interaction behaviour

(b) error rates

(c) objective search task performance

2. Related work

In the fields of computer science that seek to understand and improve users' interactions with devices - such as Human Computer Interaction (HCI), Computer Supported Collaborative Work (CSCW) and Information Retrieval (IR) - the use of various methods to capture and model these interactions is common [19]. A lot can certainly be learned from detailed examination of system-generated log data (e.g. [26, 48, 29]) or qualitative analysis of interviews and diary studies (e.g. [10]). However, such data used on its own can lack information about user context, how the user is physically holding the device and errors made when interacting with on-screen elements [36] or can suffer from inaccurate or incomplete recall of memories.

The user's surroundings can be crucial in understanding a particular behaviour, as concurrent interactions with objects other than the device of interest and distractions in the environment can have large effects [44]. This is especially true when attempting to understand interactions with mobile phones, where users often use their devices in noisy, busy and disruptive environments and contexts [24]. Recent work has called for the use of multiple video recordings, including screen recordings, wearable headcams and other sources to obtain a more complete and accurate understanding of user behaviour in context [36].

2.1. Attentional shifts

Mobile touchscreen devices have become increasingly popular, yet typing on virtual keyboards whilst walking is still an overwhelming task [40]. 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 interacting with information on the screen [39]. Bergstrom-Lehtovirta et al. [3] varied walking speed on a treadmill and measured the resulting effects on the abiulity of users to tap discreet interface elements on the touchscreen. Distractions encountered during walking can preoccupy users, reducing their effectiveness in interacting with a user interface (UI) and adding to their cognitive workload [24]. 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 [28]. As a user interacts with a mobile phone, distractions caused by walking can influence the way they hold the device, influencing the level of engagement with the task and affecting the overall effectiveness of the interaction [5, 9].

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 us to understand situational search behaviour and the influences of these attentional shifts on search.

2.2. Mobile interaction and grips

Mobile interactions are commonly achieved via touch screens upon which relatively small "soft buttons" are drawn to allow the selection of items and input text. While these buttons may be easy to accurately press in an ideal environment, such as when seated, these small and non-tactile targets may be much more difficult to interact with in other distracting situations [5]. For example, buttons in mobile UIs are often too small to comfortably press, which can result in unintended interactions [31].

Work has 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 [31]. 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 [37]. They concluded that texting whilst walking results in either a reduction in input speed (but not accuracy) or a reduction in walking pace.

The limited input modalities afforded by mobile devices have a negative effect on usability [22], a problem compounded by screen size and the device's reduced ability to present information and navigational cues [6, 8]. Small screens can easily become cluttered with information and widgets (e.g. buttons, menus, windows etc.) and this presents a difficult challenge for interface designers [6]. 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 [38].

One of the most important aspects of context in mobile computing is the way in which people hold and move with their phones, which affects how they interact with it. Nicolau and Jorge [40] investigated the effects of walking and different grip types on virtual keyboard text entry, finding that ambulatory motion had a negative effect on text entry. They also found that, although two-handed interaction improved input rate, it surprisingly did not improve accuracy through additional physical stability. More recently, Eardley et al. [17, 18] investigated the grips people tend to use when interacting with different types of device with varying screen sizes. Through a series of user studies they demonstrated that different conditions afford different grip types and that the most effective grip is dependent on the context of interaction.

While these existing studies do indicate the importance of grip in interaction with mobile touchscreen devices, they do so in very restricted sets of conditions. Users are given simple, atomic tasks to perform and the context is kept strictly static. As such we do not know how or whether users transition between grip types for different tasks and what impact situational context has on these shifts in grip. Do, for example, users employ one grip for some tasks but switch to another when the required style of interaction changes and do they shift their grips in order to compensate for distractions or environmental impediments?

2.3. Recording and annotating mobile interaction

Cameras placed in settings such as control rooms, surgeries, homes, offices, and museums are used to capture technology use in-situ [35]. Videoing interactions has been popular as attempts are made to understand and interpret the subjective qualities of user behaviour, encoding events with the help of timecodes [42]. Previous research filmed and logged user behaviour and studied preferred user interaction modalities in different contexts and for different tasks [43]. Brown et al. [7] coded events observed from video recordings on a data sheet by watching and pausing the playback tape.

As mobile devices have become more ubiquitous, recording the environment where interactions take place has become important [41] and the analysis of video allows for fine-grained analysis of interaction and activity [36]. Combining methods that record the screen and lightweight cameras to capture environmental details supports this analysis of mobile interactions [41, 7, 36]. However, while many early studies were able to assess user performance, they lacked analysis of the way that context affects interaction [4, 36].

The emergence of video annotation tools have raised the possibility of exploring not just which UI elements a user interacts with, and for how long, but also the context of use. Such tools offer the flexibility to model several perspectives over the same video content allowing multiple views of the same video data [12]. Følstad et al. [20] noted that practitioners tend to use commercially available software tools for analysis. Other research has, however, developed plug-ins for existing commercial software, for example Techsmith's Morae, bridging the gap between academic and commercial tools [27]. However, as we move towards more complex interactions, software like Morae, which is designed primarily for static lab-based evaluations, lacks the flexibility to model user interactions in-situ from these multiple perspectives.

2.4. Mobile information retrieval (MIR)

Improvements in mobile technologies have led to a dramatic change in how and when people access and use information, and have "a profound impact on how users address their daily information needs" [11]. Large-scale analysis of mobile search logs [29] 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.

Situational conditions have a number of fairly profound effects on user perceptions, both before and after completing the tasks [24], exerting a range of effects on performance and increasing difficulty of interaction [31]. This can cause users to feel more rushed, meaning they are less likely to explore the search results and to assess potentially relevant documents for relevance [13]. Tasks that are comparatively easy when stationary will likely incur a higher "cost" [1] if input accuracy [39] and reading comprehension [2] are reduced.

Environmental distractions can preoccupy users [41], reducing their effectiveness in interacting with the UI [5, 34] and resulting in a larger number of misspelled queries and an attempt by users to shorten queries [45]. Walking whilst using a mobile device requires both cognitive and motor abilities and users must divide their attention between the two tasks [32]. These different situational conditions impact search behaviour and, consequently, search performance [24, 23].

As research into information retrieval continues to evolve, evaluating search behaviour with the use of video will permit the identification of patterns and search behaviours unique to the user's condition, which log analysis can not uncover. Evaluating spatial awareness and any shifts in attention as users grip the device to type queries in these everyday situations will help to assess the levels of immersion and user abilities in mobile search tasks.

3. Method

We conducted a laboratory experiment with a total of 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 3 conditions, there was an equal spread across the two sexes (i.e. male and female), with participants from each sex assigned to all conditions. 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 (χ^2=5.13, p-value=0.74). 18

of the participants were native English speakers and the others were completely fluent in the language.

The study used a *Moto X Style* Smartphone running Android version 5 with the Google Chrome web browser for all tasks. The level of distraction and encumbrance was varied by simulating three everyday situations experienced by mobile device users: walking quickly on a treadmill, 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 per condition, and the distraction level was a between-subjects variable. Following the procedure of Lin et al. [34], participants on the treadmill were asked to select a comfortable walking pace using the increase and decrease belt speed buttons. This chosen speed was then increased by 20% to induce a small amount of ambulatory distraction. The obstacle course group was shown how to navigate a pre-defined figure-of-eight layout around tables, 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.

We developed a simple mobile search interface named *zing*, shown in Figure 1, which mimics a standard search interface by showing 10 links in descending order of relevance together with short 5-line snippets for each. The interface allows participants to enter search terms and indicate via check-boxes which documents, if any, they think are relevant. It shows the current task at the bottom of the screen, it allows participants to progress to the next search task at any time and it prompts users to fill in pre- and post-topic questionnaires. These are used to survey their perceptions about the task, their self-assessed post-task performance, satisfaction, perceived time pressure and focus/involvement on the task.

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 [49], of which we chose 4 at random from a subset of

Table 1: TREC topics used for user studies.

No.	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	3.00
404	Ireland peace talks	0.28	3.25	2.79

those that 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. Participants expressed topic difficulty by means of a 5-point Likert scale. Indexing, searching and snippet generation was provided by Apache SOLR[2].

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 3-5 documents they thought were relevant. Participant actions and behaviour were recorded by means of a GoPro camera; recording of the screen by an Android application installed on the mobile devices; for the obstacle course only, a wide-angle camera was used to record global user behaviour. The GoPro camera was worn on the head capturing hand movements, grip changes and any attentional shifts (i.e. obvious movements of the head and or body away from the device and to the environment). The touchscreen and interface was recorded using a screen recording application; this application recorded the interaction points with the device, providing insights into the interaction errors (e.g. miss typing, missing areas of the UI, etc.). For the obstacle course, the third wide-angle camera - a Sony Handycam on a tripod - captured additional mobility changes and additional behavioural changes that were not obvious from the headcam

[1] After the method of Harvey et al. [25], whereby the difficulty of a topic is determined by the average precision of its title over the document collection.

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

footage. This includes any obstacle avoidance and participants visibly slowing down during interaction.

3.1. Adobe UI plugin and video processing

As the video recordings were made by three different devices, they did not all have the same frame rate, making any immediate comparison between them impossible. To remedy this we first re-encoded all of the videos to a single, fixed frame rate of 30 frames per second. We then used Adobe Premiere to collate all three videos for each participant into a single workspace and synchronised their starting times, allowing device interactions to be assessed in context. Figure 2 shows an example of the three video streams arranged to play together synchronously.

As the footage was played back, an Adobe Premiere plugin (see Figure 3) developed for this research was used to tag the interactions. All of the footage was tagged by both researchers working as a team to reduce the possibility that any events were incorrectly tagged or missed entirely. Any disagreements regarding the appropriate tag type were resolved during the tagging process. The plugin allowed us to insert a range of coloured markers, which could be placed and annotated on the timeline. The plugin allows for the marking of both instantaneous/atomic events (i.e. those with no defined time period, such as tapping a button by mistake) and events that occur over a defined and continuous time window (e.g. entering a query using the virtual keyboard). The plugin also includes a description window allowing for additional qualitative data to support the data gathering process to be entered.

Figure 4 outlines all of the marker types we used to tag the footage, which were chosen based on the related literature and on our research aims. A marker was inserted every time the participant changed grip and at the start of each piece of footage to capture the initial grip type. As these markers encoded all of the different grips used by the participants and when they transitioned from one to another, we were able to use these atomic event markers to identify which grips were being used when every other type of tagged event occurred. Note that, although they were included in our initial table of marker types, we did not use the "Shift in speed/gait" or "Obstacle avoidance" markers in our analysis. Despite the amount of video footage available from the three different sources, these were very difficult to accurately and consistently identify.

After tagging is complete, the plugin outputs all of the resulting data in Comma-Separated Value (CSV) format for analysis. The video analysis process is labour-intensive and, due to technical reasons, video streams for some participants were incomplete or missing. As such, we present analysis of the video data for 14 out of the 24 total participants. This sub-sample is representative of the larger participant group with 8 males, a similar distribution over age ranges and an even distribution over experimental conditions: 4 baseline, 5 treadmill and 5 obstacle course.

In total, 710 individual markers were created, a rate of more than one marker for every 14 seconds of recorded footage. There was a median of 51 markers per participant over a median total interaction time of 10 minutes and 24 seconds per participant. The most frequently-occurring markers were: interaction errors (205, 28.9%), reading (121, 17.0%), attentional shifts (59, 8.3%) and typing errors (41, 5.8%).

4. Results

We simulated the everyday condition of walking when using a mobile phone by means of a treadmill and obstacle course and captured users' interactions using multiple video recordings as participants completed simple search tasks. We first present the results obtained from the analysis of the log files, which summarise the objective search performance of the participants under each of the three conditions. We then focus on the data captured by tagging user interaction behaviour (e.g. grips, shifts in grip and interaction errors) from the video data.

Note that, for brevity, in results section we will sometimes refer to the three conditions by the abbreviations B for the baseline (seated) condition, OC for the obstacle course and W for the treadmill.

4.1. Objective search performance

In order to objectively evaluate search performance, we rely on several metrics: the average number of hits (relevant documents) returned per search query; the Mean Average Precision (MAP) 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.

We also consider a number of other proxies of overall search and task performance as well as metrics such as query length and query duration - the amount of time from a search task first being presented and a query being submitted.

Table 2: Objective performance measures by condition. $*$ = sig. diff. with Obstacles; † = sig. diff. with Treadmill.

Condition	**Baseline**	**Obstacles**	**Treadmill**
Number of queries/user	13	12	14
Hits/query	3.71 $*$†	2.00	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	1.00
(Ratio relevant)	0.43	0.41	0.44
Number of query terms	3.61 $*$	3.17	3.38
Query duration	39.5s $*$†	30.5s	35.0s

Table 2 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. 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, respectively). 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 appears to have been able to more accurately choose relevant documents as they achieved the best ratio of relevant bookmarked documents. The baseline group also 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[3]). 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.

4.2. Interaction errors and corrections

Measuring errors when interacting with a device gives insight into how difficult users are finding a given device, interface or set of conditions. We measure both: typing errors, atomic mistakes when attempting to enter characters using the phone's virtual keyboard; and interaction errors, e.g. accidentally tapping on an interface element near to the desired one. Table 3 shows the average number of errors made by participants in each of the three conditions. This is averaged on a per-user basis, such that the numbers represent the average number of errors made by a user under each condition. Unsurprisingly, partici-

[3]As determined by a Wilcoxon signed-rank test.

pants in the baseline condition make fewer interaction errors than those in the other two conditions. This is likely because the other two conditions present situational impairments that exert a range of effects on performance, adding levels of difficulty as interaction with the device takes place [31].

Table 3: Errors by experimental condition (mean per user; standard deviations in brackets).

Condition	**All errors**	**Interaction**	**Typing**
Baseline	10.5 (5.2)	8.0 (5.2)	2.5 (1.9)
Obstacles	18.6 (10.4)	16.0 (9.1)	2.6 (1.8)
Treadmill	22.2 (10.3)	18.6 (8.6)	4.5 (3.1)

The baseline and obstacle course have similar rates of typing errors (2.5 and 2.6 respectively), while being on the treadmill resulted in a higher rate of 4.5. These errors are possibly caused by participants demonstrating a lack of control or an increase in cognitive workload as they interact on the treadmill [23]. On the obstacle course, however, participants can control their walking speed, giving them more control as they interact, and thus reducing error rates. Mizobuchi et al. [37] observed no reduction in input accuracy when walking and texting - the participants simply reduced their walking speed to prioritise text input. However, the walking condition on a treadmill takes away that level of control as the speed the participant must maintain is constant and this impacts on performance and difficulty, increasing the number of errors.

Interestingly, although the interaction log data indicates that baseline users spent significantly longer formulating queries, analysis of the video data shows that they spent less time actually typing those same queries using the virtual keyboard. The median query typing time for the baseline users was 10.46 seconds, while for the obstacle course and treadmill users it was 16.38 and 15.4 seconds. This is despite the fact that the baseline users' queries were significantly longer. This suggests that baseline users were able to spend more time actually thinking about the content of their queries and less time physically interacting with the virtual keyboard to enter them into the search box.

Table 4: Query corrections by condition (mean per user; standard deviations in brackets).

Condition	**Average number of corrections**	**Average duration**
Baseline	1.33 (0.58)	1.29s (1.20s)
Obstacles	2.50 (1.00)	11.58s (15.86s)
Treadmill	4.67 (2.31)	20.95s (8.82s)

Typing errors are made whilst users are inputting or amending queries and, once noticed by the user, must subsequently be corrected through further interaction with the device and, particularly, with the on-screen virtual keyboard. Table 4 shows the average number of times a user in each condition had to correct a previous mistake in typing and the cost in terms of time that each of these corrections incurred on average. As discussed above, baseline users make few typing mistakes and, therefore, have to make fewer corrections. In addition, when they do have to correct typing mistakes, it takes them far less time than users in the other conditions and, as such, they incur a much smaller overall interaction cost. The maximum amount of time for baseline users to correct a typing error was only 1.96 seconds, while for the obstacle course and treadmill conditions the maximum duration was 18.00 and 27.72 seconds respectively.

4.3. Query amendments

When searching to complete a given task it is often necessary to amend the original query. This is often done after users have assessed the results of the initial query and, based on this, either choose additional keywords to add to the original query or to remove extraneous keywords from it. Research shows that query amendment/refinement generally leads to better overall search performance and that it is a beneficial technique often employed by expert users [50].

As outlined in Table 5 all participants in the baseline condition made some query amendments: at least 4 instances per user with a median duration of 8.42s. However, only two treadmill users and three obstacle course users made

Table 5: Query amendment statistics.

Condition	Instances	Median time	Total time	Number of users
Baseline	22	8.4s	225.8s	4 (all)
Obstacles	5	8.9s	46.8s	3/5
Treadmill	9	10.6s	118.8s	2/5

any amendments, although these tended to take more time to complete.

4.4. Attentional shifts

Shifts in attention are defined as either *major* or *minor* within this study. Major shifts are a clear movement of the head away from the device, while minor shifts are a small movement in which the participant's gaze is still set within the UI but there has been a movement that could be seen as a slight distraction. These attentional shifts were assessed via careful analysis of the head-mounted GoPro footage. These may be caused by attention being drawn away from the interaction by external distractions in the surrounding environment or, in some cases, by distractions local to the user, for example needing to scratch one's nose.

Table 6: Attentional shifts by condition (average count per user; standard deviation in brackets).

Condition	All shifts	Minor	Major
Baseline	3.75 (2.99)	3.50 (3.11)	1.00 (0.00)
Obstacles	2.25 (1.89)	2.00 (0.00)	1.67 (1.15)
Treadmill	7.00 (10.17)	7.50 (9.11)	2.50 (2.12)

Table 6 shows the major and minor attentional shifts identified by experimental condition. The results show there are considerably more minor attentional shifts in the treadmill condition compared to the other two conditions, although the standard deviations are high. Interestingly, participants on the obstacle course had the smallest number of shifts overall and, based on the video observations, this condition showed that participants are more immersed within the activity and are concentrating on their route around the course. There was only one instance where a participant on the obstacle course needed to avoid obstacles in a manner that could be deemed to have clearly affected their behaviour.

Table 7: Average and total duration of attentional shifts by condition (all in seconds; standard deviations in brackets). B = baseline, OC = obstacle course, W = treadmill.

	All shifts		Minor		Major	
Condition	**Total**	**Ave.**	**Total**	**Ave.**	**Total**	**Ave.**
B	61.56	4.10 (3.02)	58.60	4.19 (3.12)	2.96	2.96 (0.00)
OC	30.72	3.41 (2.63)	17.56	4.39 (3.71)	13.16	2.63 (1.34)
W	110.06	3.14 (2.13)	101.89	3.40 (2.20)	8.17	1.63 (0.57)

The actual duration of the shifts in attention (see Table 7) tend to be more or less the same by condition, although they are somewhat shorter on average in the baseline condition and there were very few instances of major attentional shifts. This is perhaps due to the controlled environment within the lab lacking the realism of true mobile interactions.

4.5. Reading

Table 8: Median number of documents read and reading time by condition. B = baseline, OC = obstacle course, W = treadmill.

	B	**OC**	**W**
Documents read per user (count)	15.0	6.5	8.0
Reading time per document (seconds)	13.2	14.1	13.1
Total reading duration per user (seconds)	201.6	92.1	105.4

Another important behaviour for good search performance is the amount that people read the content of search results - the more time spent reading, the more likely it is that a searcher will be able to identify good quality, relevant documents [50]. Table 8 displays various statistics relating to reading behaviour over the three experimental conditions.

People on the baseline read considerably more documents and read for a much longer amount of time overall; however, the amount of time spent reading an individual document is much the same over all the conditions. These results are in contrast to Barnard et al. [2], who found that participants spent more time reading the passages and trying to encode them in the walking condition, likely because the process of encoding was hindered by their motion.

4.6. Grips

An important factor when studying how people interact with objects is the grip that they choose to use. This is especially important for interaction with mobile devices and can be influenced by a number of contextual factors [18]. When tagging the video footage, we identified each time a participant changed their grip between 1 of 5 different grip types. Four of these (grips A–D) are after the work of Eardley et al. [18] and are shown in Figure 5. The fifth (grip E) was only used in the baseline condition and refers to the situation in which the user interacts with the device when it is laid on the table.

Figure 6 shows the number of times participants switched to each of the grips by condition, expressed as a percentage of the total number of shifts in grip made in that condition. We observe that there are considerable differences in the use of the various grips depending on the condition. Overall, we observed 53 shifts in grip by the baseline users, 56 by the treadmill users and only 9 on the obstacle course.

On the obstacle course D is the preferred grip type, while on the treadmill the preference seems to be for grip type B. In the baseline condition the most frequently occurring grip is A, although there appears to be a much more even distribution in this context. Note that grip A is not used at all on the obstacle course and grip B is only used once, while on the treadmill grip D is only used twice.

It may be possible that, even though a given grip is only used a small number of times in a given condition, its duration of use still constitutes a large proportion of the total interaction time. Figure 7 shows the total duration of time that each grip was used under each of the three conditions. Here we see some slight differences when compared to the counts in Figure 6. Although grip D is only used twice by treadmill users, it actually accounts for nearly 40% of the total duration. Further analysis showed that the two users on the treadmill who used this grip did so for the entire duration of the study. Grip A was the most frequently occurring grip for baseline users, however, when duration is accounted for, it becomes the second most used after grip D. Grip C is used quite infrequently and, when it is used, is only maintained for a very short period of time.

4.6.1. Errors by grip type

We now consider the number of errors made by participants when using each of the 4 main grip types. To do this we identified from the tagging data the grip that was being used when each tagged interaction was made, allowing us to investigate errors by grip. Of the 246 errors identified in total (i.e. both interaction and typing errors), 12.6% were using grip A, 22.8% when using B and nearly two-thirds (63.8%) were made when using grip D. In contrast, only 2 errors occurred while participants were using grip type C. We can normalise these numbers by considering the percentage of all interactions conducted using each grip that were errors. Again, grip D appears to be the least stable, with nearly half (49.4%) of interactions using this grip being errors, while grips A and B seem to be somewhat more stable, resulting in errors in only 27.4 and 29.9% of interactions respectively.

Figure 8 shows the ratio of interactions using each type of grip that were errors, segmented by experimental condition. In other words, for a given experimental condition, when users were employing a given grip type, what ratio of their interactions were recorded as errors. Although grip D is used for large amounts of time in all 3 conditions, the error rate is significantly [4] lower in the baseline condition (16.2%) than on the treadmill (62.6%) or the obstacle course (57.2%). Grip B, on the other hand, only results in high error rates for the obstacle course condition (47.6%). Grip A was only observed in use by baseline users and those on the treadmill and caused slightly higher error rates on the latter (34%) than the former (22.2%).

5. Discussion

This research set out to investigate searching on the go and the effects common mobile situations have on hand movements, grip and shifts in grip, assessing the impact of context and grip on interaction behaviour, error rates and search task performance. The following discussion will consider each of these areas of interest in line with the aims outlined in the introduction.

The effects mobile situations have on hand movements, grip and shifts in grip.

The results of the study demonstrate a number of effects on interaction and user behaviour. The outcomes highlight substantial variations between situations and both the number of changes in grip (grip shifts) and the frequency of use of each grip type. Overall, across all three conditions, participants tended to primarily use two-handed grips, with only very little use of grip C (the sole one-handed grip). This is perhaps not surprising as a two-handed grip should provide a more stable base for interaction, especially if the participant has to maintain balance when walking. Baseline users tended to use all of the grip types and were much less likely to stick to a single grip for large amounts of time. This may be because this is a less distracting setting and grip changes may have occurred because subjects felt more relaxed [23] and free to make them. Research suggests that the baseline users feel less rushed [24] and may

[4]Significance determined from Z scores of 2 (i.e. pairwise) population proportions. $p \ll 0.01$ for both comparisons.

feel that they have more time to construct queries [30] and use different grips, allowing them to alter interaction behaviour depending on task requirement. For example, baseline users often employed grip A when typing queries as it permits more rapid typing but then switched to grips B or D when interacting with search results or reading. They were also the only group to use grip C for any considerable length of time, likely because the lack of ambulatory movement means that it is not necessary to use both hands to stabilise the device.

The treadmill had the largest number of grip shifts, a factor possibly influenced by the extra cognitive effort needed to complete the tasks whilst walking [37], especially at speeds above a subject's normal pace. In comparison with those on the obstacle course, participants on the treadmill do not have control over how quickly they are walking and cannot simply reduce their pace when reaching a cognitively challenging part of the task. Earlier research [23] noted the treadmill to be the most difficult due to the lack of control over speed whilst walking. Initial insights suggest that grip shifts under these conditions reflect an increased cognitive load for subjects, which leads to behavioural modification as evinced by the work of Azzopardi et al. [1]. In contrast, the obstacle course group only very infrequently changed their grip and tended to maintain the same grip over long periods of time, even if the tasks they were performing changed. Unlike in the other two conditions, they often did not change grips when transitioning from text input to browsing or scrolling. Walking and avoiding obstacles necessitates the user's brain frequently switching attention between the environment and the device [16] and the additional requirement of a change in grip type may perhaps be of lesser priority, even though it might improve interaction performance.

The impact of context and grip on interaction behaviour, error rates and search task performance.

Situational context clearly had a considerable impact on search performance - those in the baseline condition were able to achieve significantly higher MAP scores and almost double the average number of hits per query than those in the other two conditions. Some reasons for this large difference were revealed through analysis of the interaction logs: baseline users submitted significantly longer queries and spent more time actually generating queries. Analysis of the video footage, however, showed that the baseline users actually spent less time typing their queries into the search box and, therefore, were able to spend more time selecting appropriate keywords.

Amendments made to queries and reading of result-set documents both indicate a high level of engagement with the task and are behaviours commonly employed by expert searchers [50]. In total over all conditions we identified 36 instances of query amendment, 22 of which were made by the baselines user, all of whom were responsible for at least four. This was not the case in the two walking conditions as only two treadmill and three obstacle course users made any amendments at all.

We observed a similar behaviour when it came to reading documents to assess relevance: baseline users read frequently and, over the entire study, for a long period of time, while those in the other conditions read much less frequently. These differences in behaviour suggest that the walking conditions require more effort from users to perform these tasks [41] and, as such, they engage in them as little as possible. This notable lack of exploratory search behaviours and under-prioritisation of such tasks has been identified before when artificial time constraints were imposed on searchers [9, 14]. Although we imposed no such restrictions, the walking conditions may have nevertheless made users feel rushed and under pressure [24].

Although both typing and interface interaction errors were observed under all conditions, there was considerable variance in the incidence of errors by condition. Perhaps unsurprisingly, the baseline condition yielded the fewest errors overall, particularly interaction errors, which were much more frequent in the other two conditions. This is most likely because there is no body movement for baseline users to compensate for and, as such, they are much less likely to miss small targets when tapping [26, 3, 41]. Typing errors were far more common in the treadmill condition than either of the other conditions. This is likely also related to the inability of users in this condition to reduce their pace when encountering a cognitively challenging task, unlike those on the obstacle course, who often did [37, 5]. These typing errors then necessitate later correction of the incorrectly typed word(s), a task which the baseline users were able to perform much more quickly, incurring much smaller interaction costs for their earlier mistakes.

Incidence of errors also varied by combination of grip type and condition. Interestingly, when baseline users interacted using grip D the error rate was very low (lower than with either grips A or B), while for the treadmill and obstacle course conditions this was certainly not the case as the error rate when using grip D was the highest for both conditions, especially for the treadmill. This suggests that, although grip D seems to be commonly used when walking and may feel intuitive under such conditions, it may not actually be a good choice. This may be because the fingers that are interacting with the screen are on a hand that is not actually cradling the device and is completely free to move and, therefore, may easily be jostled by ambulatory movement or when avoiding obstacles. This is in contrast to grips A an B, where the “interacting hand” is somewhat braced against the device at the palms and is thus more “anchored” and stable. The baseline results agree with those of Eardley et al. [18], whose participants found grip D to be the most secure and comfortable. All experiments in Eardley et al. [18] were conducted with users seated and our results show that their conclusions may not hold when users are in other situational contexts.

6. Conclusions

By simulating common mobile contexts and analysing both log data and interaction data obtained by tagging video footage from three different sources, we were able to repeatably investigate user behaviour, search performance and error rates under different experimental conditions. This combination of analysis methods allowed us to further identify the grips participants were using as they completed tasks, what effects these had on their interactions and how these effects resulted in interaction and typing errors.

Using both sources of information we revealed considerable differences in performance and behaviour between the three conditions, demonstrating the effects that the context in which mobile interaction takes place has. Detailed analysis of the video footage allowed deeper insights into search behaviour and interaction that would not have been possible with the log data alone. Our results provide useful insights to inform the design of future mobile search interfaces, giving us a greater understanding of how situational contexts, such as walking, impact search performance and user behaviour. Our results also build on previous work on grips (e.g. [17, 18]) to provide novel insights into how grip use varies under different mobile conditions and how these conditions affect the effectiveness of these different grips.

For future research in this area, we plan to expand the scope of this work by considering other more natural/field-based contexts, such as a busy high street, public transport or a bar. While simulation of walking and navigating obstacles in a lab gave us control over many aspects of the conditions, they do not fully simulate the difficulties encountered when interacting with a mobile device whilst walking. The obstacles on the course were static and participant movement was repeated and predictable; what impact do unpredictable objects and obstacles have on user behaviour and how do users employ grips to counteract this? We intend to design search interfaces that adapt to walking using phone accelerometers to investigate whether a UI can respond to situational changes to improve user experience and reduce interaction errors. We also intend to investigate the possibilities of a mobile phone detecting how it is being held and responding to particular grip types.

References

[1] Leif Azzopardi, Diane Kelly, and Kathy Brennan. How query cost affects search behavior. In *Proceedings of the 36th international ACM SIGIR con-*

ference on Research and development in information retrieval, pages 23–32. ACM, 2013.

[2] 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(2):81–96, 2007.

[3] Joanna Bergstrom-Lehtovirta, Antti Oulasvirta, and Stephen Brewster. The effects of walking speed on target acquisition on a touchscreen interface. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 143–146. ACM, 2011.

[4] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*, pages 47–56. ACM, 2011.

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

[6] Stephen Brewster. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188–205, 2002.

[7] Barry Brown, Moira McGregor, and Donald McMillan. 100 days of iphone use: understanding the details of mobile device use. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 223–232. ACM, 2014.

[8] 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(3):165–181, 2004.

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

[10] Karen Church, Antony Cousin, and Nuria Oliver. I wanted to settle a bet!: understanding why and how people use mobile search in social settings. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 393–402. ACM, 2012.

[11] 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(2):10, 2014.

[12] Miguel Costa, Nuno Correia, and Nuno Guimarães. Annotations as multiple perspectives of video content. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 283–286. ACM, 2002.

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

[14] Anita Crescenzi, Diane Kelly, and Leif Azzopardi. Impacts of time constraints and system delays on user experience. In *ACM CHI*, pages 141–150. ACM, 2016.

[15] Jerry Dischler. Building for the next moment. *Google Inside AdWords*, May 2015, 2015.

[16] Mark Dunlop and Stephen Brewster. The challenge of mobile devices for human computer interaction. *Personal and ubiquitous computing*, 6(4): 235–236, 2002.

[17] Rachel Eardley, Steve Gill, Anne Roudaut, Stephen Thompson, and Joanna Hare. Investigating how the hand interacts with different mobile phones. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 698–705. ACM, 2016.

[18] Rachel Eardley, Anne Roudaut, Steve Gill, and Stephen J Thompson. Understanding grip shifts: How form factors impact hand movements on mobile phones. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4680–4691. ACM, 2017.

[19] Gerhard Fischer. User modeling in human–computer interaction. *User modeling and user-adapted interaction*, 11(1-2):65–86, 2001.

[20] Asbjørn Følstad, Effie Law, and Kasper Hornbæk. Analysis in practical usability evaluation: a survey study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2127–2136. ACM, 2012.

[21] PA Hancock, BD Sawyer, and S Stafford. The effects of display size on performance. *Ergonomics*, 58(3):337–354, 2015.

[22] 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(1):1, 2013.

[23] Morgan Harvey and Matthew Pointon. Perceptions of the effect of fragmented attention on mobile web search tasks. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 293–296. ACM, 2017.

[24] Morgan Harvey and Matthew Pointon. Searching on the go: the effects of fragmented attention on mobile web search tasks. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 155–164, New York, NY, USA,

2017. ACM. ISBN 978-1-4503-5022-8. doi: 10.1145/3077136.3080770. URL `http://doi.acm.org/10.1145/3077136.3080770`.

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

[26] Niels Henze, Enrico Rukzio, and Susanne Boll. 100,000,000 taps: analysis and improvement of touch performance in the large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 133–142. ACM, 2011.

[27] Jonathan Howarth, Tonya Smith-Jackson, and Rex Hartson. Supporting novice usability practitioners with usability engineering tools. *International Journal of Human-Computer Studies*, 67(6):533–549, 2009.

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

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

[30] Maryam Kamvar and Shumeet Baluja. Deciphering trends in mobile search. *IEEE Computer*, 40(8):58–62, 2007.

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

[32] Eric M Lamberg and Lisa M Muratori. Cell phones change the way we walk. *Gait & posture*, 35(4):688–690, 2012.

[33] 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(7):e0133281, 2015.

[34] 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(9):759–769, 2007.

[35] Paul Luff and Christian Heath. Mobility in collaboration. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 305–314. ACM, 1998.

[36] Donald McMillan, Moira McGregor, and Barry Brown. From in the wild to in vivo: Video analysis of mobile device use. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 494–503. ACM, 2015.

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

[38] 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. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, pages 427–436, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3673-4. doi: 10.1145/2838739.2838748. URL `http://doi.acm.org/10.1145/2838739.2838748`.

[39] Matei Negulescu, Jaime Ruiz, Yang Li, and Edward Lank. Tap, swipe, or move: attentional demands for distracted smartphone input. In *Proceedings*

of the International Working Conference on Advanced Visual Interfaces, pages 173–180. ACM, 2012.

[40] Hugo Nicolau and Joaquim Jorge. Touch typing using thumbs: understanding the effect of mobility and hand posture. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2683–2686. ACM, 2012.

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

[42] Antti Pirhonen, Stephen Brewster, and Christopher Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 291–298. ACM, 2002.

[43] Tiago Reis, Marco de Sá, and Luís Carriço. Multimodal interaction: Real context studies on mobile digital artefacts. In *International Workshop on Haptic and Audio Interaction Design*, pages 60–69. Springer, 2008.

[44] Simon Robinson, Gary Marsden, and Matt Jones. *There's not an app for that: mobile user experience design for life.* Morgan Kaufmann, 2014.

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

[46] Richard Schaller, Morgan Harvey, and David Elsweiler. Out and about on museums night: Investigating mobile search behaviour for leisure events. In *Searching4Fun WS at ECIR*, 2012.

[47] 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(6):535–541, 2016.

[48] Jaime Teevan, Eytan Adar, Rosie Jones, and Michael AS Potts. Information re-retrieval: repeat queries in yahoo's logs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 151–158. ACM, 2007.

[49] E.M. Voorhees. The trec 2005 robust track. *SIGIR Forum*, 40(1):41–48, June 2006. ISSN 0163-5840. doi: 10.1145/1147197.1147205. URL http://doi.acm.org/10.1145/1147197.1147205.

[50] Ryen W White, Susan T Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*, pages 132–141. ACM, 2009.

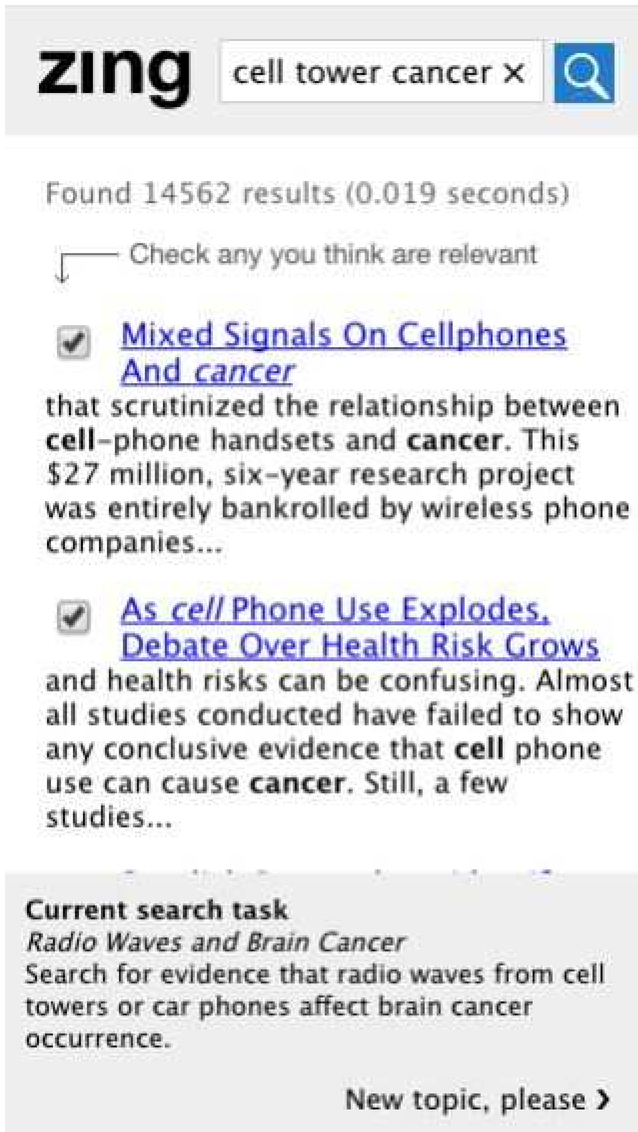

Figure 1: Zing search interface on an Apple iPhone 5. Checkboxes used to indicate relevance.

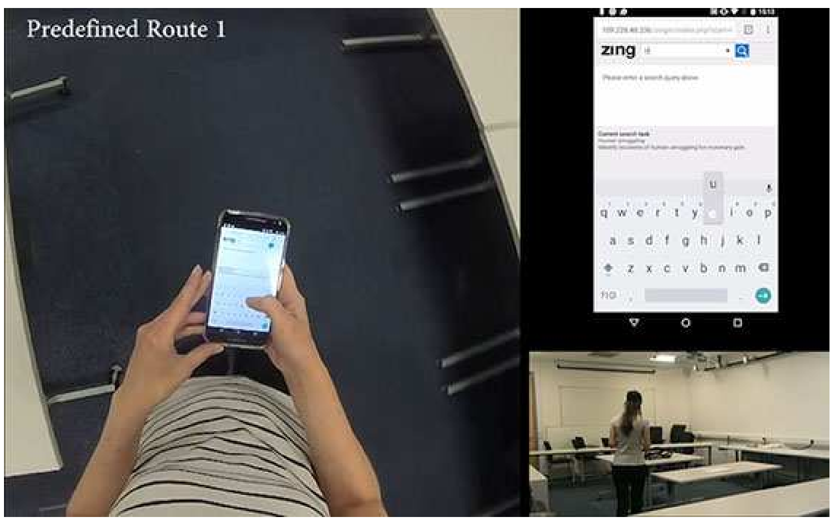

Figure 2: An example of a merged video screen.

Figure 3: Premiere plugin user interface.

Data Capture:	Marker Type	Description
Grip type	Mobility	Type and change of grip i.e., one handed or two hand. Symmetric bimanual (A) Asymmetric bimanual with the thumb (B) Single-handed (C) Asymmetrical bimanual with the finger (D)
Interaction	Error	Atomic errors made within the UI.
Typing	Error	Type of error made i.e. letter deletion, word deletion, miss typing and miss spelling queries. *[time window/atomic]*
Query generation	Instruction	Length of time to generate query. *[time window]*
Query amendment	Operator	Adding/removing query terms. *[time window]*
Query correction	Operator	Fixing previous typing error. *[time window]*
Reading time	Info-Proc	Time taken to read and process information. *[time window]*
Shift in speed/gait	Physical space	
Obstacle avoidance	Mobility	*[time window]*
Attentional shift	Physical Space	Attentional shift – considerable head movement or minor head movement. *[time window]*

Figure 4: List of marker types used for video tagging.

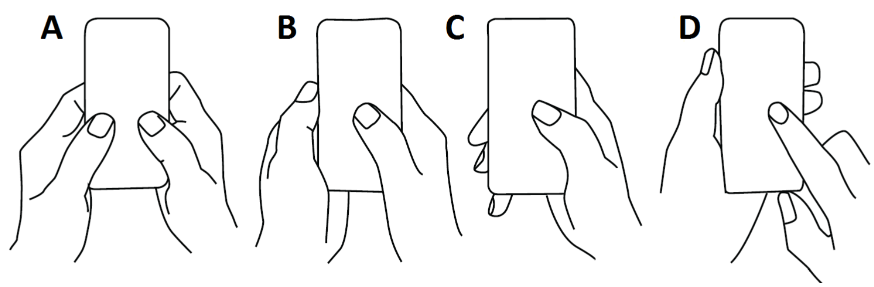

Figure 5: Visualisations of the 4 most-commonly identified grip types, after Eardley et al. [18].

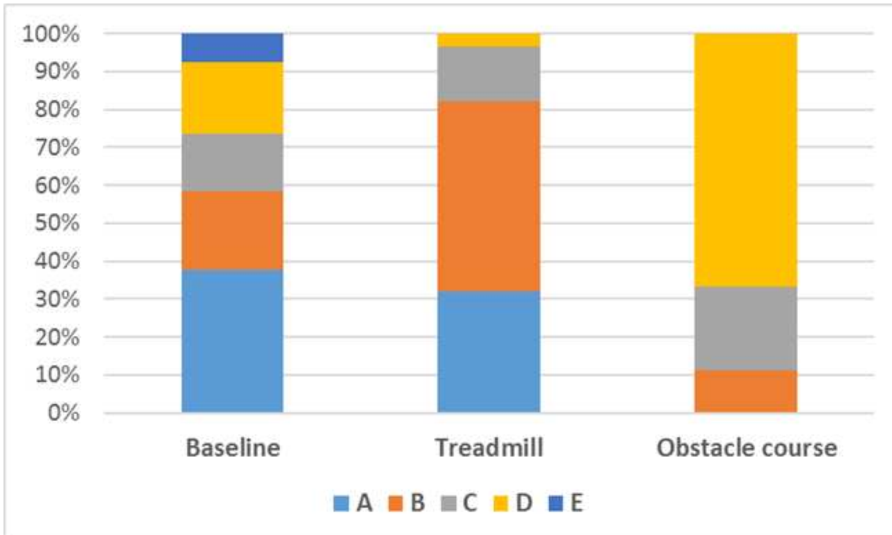

Figure 6: Grip count by condition.

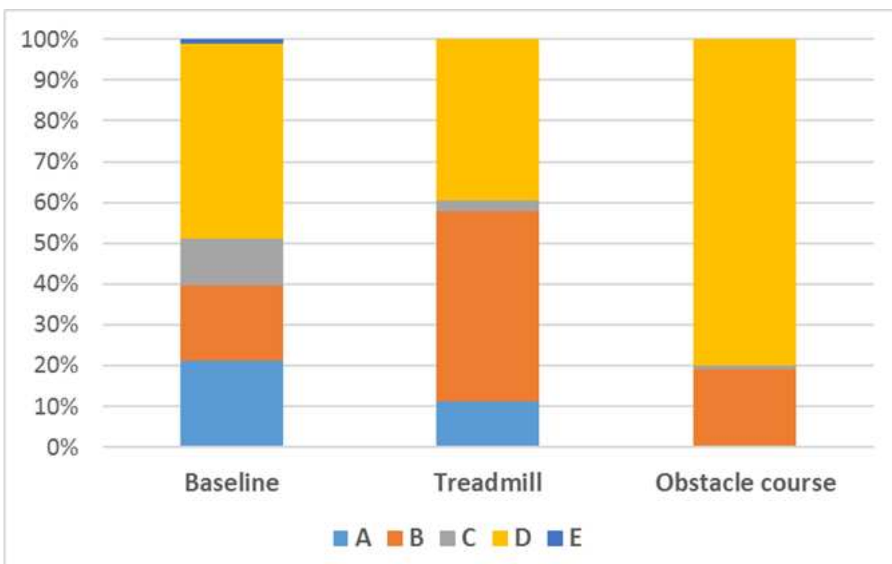

Figure 7: Grip time by condition.

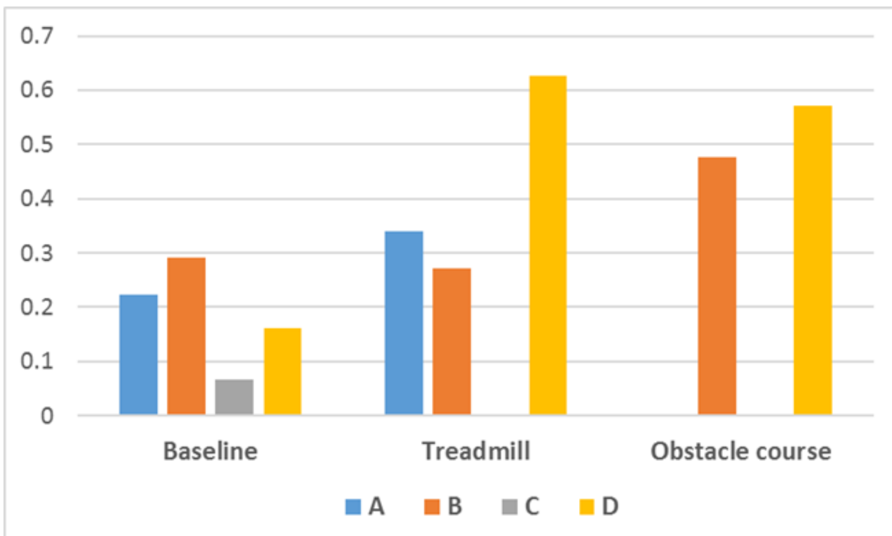

Figure 8: Errors by condition as ratio of tagged interactions using each grip type.