ABSTRACT

Festivals held in a city (or number of cities) contain many geographically distributed events often occurring contemporaneously. Visitors to the festival need to make numerous cognitively-challenging decisions about which events to see, and in which order. Consequently, the visitors’ interaction with information systems is likely to influence their experience of the festival. In this paper we investigate how such interactions with a mobile app, designed to provide visitors with information about the festival and to help them plan their itinerary, relate to their experience and how they participated in the festival. The app was deployed in a large-scale naturalistic study (n=1159). Our analysis reveals that different information interaction styles corresponded to itineraries with different properties. The results of a follow-up survey (n=59), completed by a subsample of these users, suggests that this is no coincidence. Analysing what people reported in terms of their preferences for their evening reveals trends indicating that user groups who made use of the same interface features (i.e., search, browse or recommendation) had similar priorities when planning their evening and ended up visiting events that reflect those priorities. These findings suggest that users are able to adapt their interaction style to use the features most appropriate to their needs. We conclude by discussing what our findings mean in terms of the information behaviour literature and evaluating interactive information retrieval systems embedded in a real context.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval: Search Process; H.5.m. [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous

Keywords
Mobile Assistance System, Distributed Events, Information Needs, Information Behaviours

1. INTRODUCTION

A distributed event is a collection of smaller, single events occurring at around the same time and conforming to one overarching theme. Well known examples include the Edinburgh Festival Fringe and the Montreal International Jazz Festival. What many of these events have in common is a huge number of diverse geographically and temporally dispersed sub-events. For example, in 2011 the Edinburgh Fringe had over 2,500 shows at 258 venues, ranging from the classics of ancient Greece to contemporary works.

While variety is a major selling point of such events, making the decision of which sub-events to visit is not trivial; visitors have to discover which are on offer and select a small subset from the many they find appealing. The decision depends on factors such as user preferences, time constraints, the location of events and available transport connections. People often feel overloaded with information and rely on tips from friends and on serendipitous discovery [16]. Without careful planning this could mean missing events of interest or spending large amounts of time travelling. Information systems, such as mobile phone applications [16], web based portals or recommender systems [17] present a clear opportunity to assist users in performing this task.

The focus of this paper lies in behaviour when using such assistance systems. We investigate how visitors to a popular distributed event (the Long Night of Music) interact with a rich mobile application designed to support them in finding events they will like and combine these into a plan for the evening. In particular we explore the relationship between user behaviour, in terms of the features of the system they interact with, and the outcome of their evening e.g. the number of different events they visit, where those events are located and the amount of time spent travelling. We show that different information interactions (or styles of interaction) lead to distinctly different outcomes: specifically the user’s itinerary. Therefore their experience of such an event relates to whether they prefer to search, browse or receive recommendations.

2. RELATED WORK

2.1 Human Information Behaviour

The Human Information Behaviour (HIB) literature shows that behaviour with information systems can vary considerably depending on a number of different user, system and contextual factors [9]. Users are often grouped into types based on their behaviour, facilitating comparison and insight into how these behaviours manifest. Two groups of
users whose behaviour has frequently been compared are novices and experts [2, 24]. For example, in the context of web search, distinctly different behaviours have been observed which lead to different task outcomes: experts spend less time generating queries, take less time clicking on results, explore results more thoroughly; and are ultimately more successful. However, generally users will tend to issue short queries and only examine the top of the results list [22]. This is despite the fact that those who are willing to submit longer queries are often rewarded with more effective results [10]. In addition users would further benefit if they examined more of the results list, yet seldom do so [11]. One interpretation for such findings is that this approach is simply not economical [4]. In [12] it is shown that users could still be effective in finding relevant information by posing a series of extremely short (and cheap) queries. This suggests that users evolve behaviours that let them get the most of the system for the least effort.

In the HIB literature, a feature of research performed to date is that investigations take place in the context of one particular type of system e.g. search or recommender systems. The presented work addresses both of these issues by examining more of the results list, yet seldom do so [11]. One interpretation for such findings is that this approach is simply not economical [4]. In [12] it is shown that users could still be effective in finding relevant information by posing a series of extremely short (and cheap) queries. This suggests that users evolve behaviours that let them get the most of the system for the least effort.

Contextual changes can also influence behaviour. [3] shows that when search tasks get more difficult, users tend to formulate more diverse queries that feature more advanced operators and spend more time on results pages. People adapt their strategies to the capabilities of the system e.g. when search systems are less effective users respond by submitting more queries or looking deeper in the results list [12, 21]. In [21], the authors showed that users adapted to a degraded search system by posing more queries.

Cognitive biases can play a role in shaping the search behaviour of users. For example, it was shown that people tend to prefer positive information, i.e. an affirmation bias [23], and that they look for information that supports things they already know or that conform to existing viewpoints, i.e. confirmation bias [20].

Trust has been shown to influence how users will interact with a system. In the context of recommendations, it has been shown that it is important for the system to build up a level of trust with the system, than simply trying to improve the accuracy of the recommendations [7]. It has also been demonstrated that recommender systems can exert a significant influence on the choices and evaluations users make and even on their perception of items [6].

The reviewed literature demonstrates that behaviour with information systems can vary and this can relate to numerous factors including the user’s innate abilities or experience with the system, the effort he is willing to put in and underlying personal biases. However, other factors can be important too. Different users tend to naturally prefer particular strategies of usage and can tailor these to particular tasks and contexts. Finally, there is strong evidence that the behaviour exhibited influences the outcome or success of the task. While most of the literature focusses on traditional, work-oriented, search tasks and work goals, in this paper we perform one of the first studies that explores how users adapt their behaviour in a leisure information seeking context.

2.2 Tourism and Distributed Event Research

Tourists visiting a city face similar problems to those of distributed event visitors, i.e., finding interesting sights to visit, deciding which to actually visit, planning itineraries, etc. Various electronic tourist guides have been developed (e.g. [8, 13]) to assist these visitors by offering recommendations, generating routes and supporting users enroute, which is similar to the features of the app discussed in this paper. Limited evaluation of such systems restricts the input we can draw for our work. However, one important discovery was Kramer et al’s finding that “tourists visit a lot of unplanned attractions, likely for two reasons: they stop at additional attractions when passing by along their normal route or they explicitly deviate from the given itinerary to visit a certain attraction they saw from afar.”

Another, somewhat smaller, body of literature has investigated what users want and how they behave to achieve what they want during distributed events. [15] interviewed visitors of two such events to find out what they wanted from the experience. The results show a multitude of desires from wanting to visit a specific event or events to do with a particular topic or genre of music, to wishing to discover and visit a range of novel and diverse events. Some participants reported wanting to enjoy time with friends or family, to learn or simply to escape the television. Other aims were more practical, such as the desire to limit the amount of travel time or to stay within a particular area.

Other work has tried to understand how users interact with information systems to achieve these aims. [16] investigated behaviour with a search interface during a distributed event, revealing that users mostly typed very short queries, composed primarily of named entities, suggesting that the system was used as a means to filter for events they already knew about, rather than to discover events with particular properties or with particular themes.

Recommender systems are another means of providing user support in tourism or distributed event settings. [17] investigated a number of recommendation algorithms in the context of distributed events and found that subtle differences in how the systems worked resulted in quite different user behaviour. The authors concede that while the recommendations generally helped users to select interesting events and build good itineraries, different groups of users had different objectives and might be better served by a variety of options, rather than a single catch-all solution.

More closely related to this work, [15] presented a mobile application to support users at two popular distributed events. Via analysis of user interaction with the app it was shown that participants who used different features of the system tended to rate events with different properties as interesting. The authors believe this suggests that different groups of users had quite different aims in mind. However the authors did not ascertain whether or not these events were actually visited by the users which is important as people tend to mark many times more events as interesting than they ultimately choose to visit.

Within distributed events what is lacking is an understanding of how different behavioural strategies and system support relate to outcomes. Evaluating search outcomes often relates to success i.e. being able to perform the task that motivated the information need in the first place, but in other scenarios different criteria for success are useful i.e. do distributed event visitors achieve the desires reported in the interviews in [15]?

In the HIB literature, a feature of research performed to date is that investigations take place in the context of one particular type of system e.g. search or recommender systems. The presented work addresses both of these issues by answering the following research questions:

- RQ1: Are there preferences for types of system (search,
by choosing one of the six tabs depicted in Figure 1: 

- RQ2: If yes, do different modes of interaction lead to different outcomes for users i.e. do the events they visit have different properties?
- RQ3: If yes, are these properties desired i.e. do they match what users want from the evening?

Addressing these questions not only provides insight into how well our system is performing, but our analyses allow us to better understand the relationships between intention, behaviour and outcomes - something scarcely mentioned in the current literature.

3. THE LONG NIGHT OF MUSIC 2013

In this paper we investigate behaviour during the Long Night of Music, an annual cultural event organised in the city of Munich. In addition to a diverse range of age groups and social backgrounds. In 2013 approximately 20,000 people visited a total of 212 events at 113 distinct locations including live performances, concerts, dance classes, multimedia shows and exhibitions. The music at the events ranged from classical to hip-hop to death metal. The events were dispersed across the city and its suburbs and designated bus tours were set up to transport visitors between events.

Events can be discovered by means of the booklet that is distributed for free by the organisers and contains descriptions of all events in the order they lie along the bus tours. This booklet is necessarily large (110 A6 pages) and can be difficult to navigate, making it time-consuming to find events with certain criteria. In the booklet the event descriptions are ordered geographically by bus route, rather than by musical genre, and choosing the best events to visit often involves scanning the complete booklet. As shown in [15], in such contexts information systems can lessen the burden on the user, assist in finding events that they will enjoy visiting and help construct a suitable itinerary for the chosen events.

4. MOBILE APPLICATION

For the Long Night of Music we developed an Android app to help visitors achieve the aims reported in [15]. The app was designed to assist visitors in searching and browsing through the events and also recommended events based on their interests and location. The system also helped to create an itinerary for visitors which incorporated constraints like the start and end time of events, the travel time between event locations based on the public transport routes and schedules so that visitors could maximise what they could see and minimise travel time. The app also let users manually customise plans by adding, removing and re-ordering events. Based on the created plan, the application can guide the user between chosen events using a map display and textual instructions. Figure 1 shows screenshots of the app.

The user has five ways to find events he would like to visit by choosing one of the six tabs depicted in Figure 1: 

1 http://www.muenchner.de/musiknacht/

The first tab (Recommender) offers personalised event recommendations to the user, generated from a hybrid recommender as proposed by [17] which combines content-based algorithms, collaborative filtering algorithms and a temporal contiguity recommender designed to recommend event next to the already selected events.

The second tab (by Tour) allows the user to browse events organised geographically by their position on bus routes laid out by the organisers, in the same order as they are printed in the free booklet. This tab supports users in finding events based on location and may provide one way of minimising travel time between events.

The third tab (by Genre) provides a means to browse events by music genre. The organisers provided us with a music genre categorisation for each event. Users choose one of these genres from a drop-down menu and are presented with a list of all appropriate events.

The forth tab (by Search) provides free-text search functionality over the names and descriptions of the events. The event descriptions and titles are tokenised and stemmed. In order to match topically similar words we then map every token to one or more topic groups, whereby ensuring that terms such as “dinner” and “food” are mapped to the same groups. The search tab, therefore, offers support to find events by topic, type of event, event description or the title of the event itself. To speed up interaction with the system, queries were submitted after each typed character (search-as-you-type).

The fifth tab (Map) depicts a map of Munich with the event locations highlighted, giving users an overview of the spatial relation between events. This might allow users to select a number of events close to each other to minimise travelling time.

The last tab lists the events the user has already selected.

Figure 1: The by genre screen with genre Rock selected (left) and the map screen with the planned route (right)

In the remainder of the paper, in line with the research aims as outlined above, we focus on the way users made use of the app - especially how they found events of interest to them - and how these usage patterns related to their visiting behaviour on the night.
5. DATA COLLECTION

We address our research questions by examining three sources of data collected: user interaction logs with our app; GPS tracks, which allow the ascertaining of visited events and subsequently metrics for evening outcomes matching the desires reported in [15]; and questionnaire data, revealing individual user priorities for the evening.

5.1 Log data from the app

We examined user behaviour by recording all user interactions and positional data within the mobile app. The app was on offer for download from the Google Play Store and advertised on the official Long Night of Music web page. In total it was downloaded approximately 1300 times and 1159 users allowed us to record their interaction data. We recorded all interactions with the application including tab changes, tour/genre selection, click-throughs, tours generated, modifications to tours as well as all the ratings submitted for events. GPS positional updates were recorded if users turned GPS on. We received GPS data from 180 users between 5pm and 5am on the day of the events.

A short questionnaire run upon first start-up of the app provided us with demographic information. 59.7% of the app users were first time visitors to the Long Nights, 16.8% were second time visitors, 11.8% were third time visitors and 11.8% had attended more than three times previously. 0.9% of users were 17 years of age or younger, 41.6% were between 18 and 29, 30.4% 30-39, 17.4% 40-49, 7.8% 50-59 and 1.9% above 60 years old. These demographics are similar to those reported by the event organisers for a previous Long Night of Music in Munich [1] and suggest that our sample of users should reflect the visitors as a whole.

5.2 Online survey

Upon start-up of the app we ask our users to voluntarily enter their email address to take part in an online survey, which they were asked to complete after the Long Night was over. 59 users took part in this survey, which provided information on what users hoped to achieve from the evening, their behaviour with the app, the events they visited and their subjective opinions regarding the app. Although all data were anonymous, survey responses and collected log data were linked together.

6. GAUGING BEHAVIOURAL OUTCOMES

6.1 Metrics for Understanding User Experience

In order to understand the outcomes of visitor behaviour (i.e. the properties of events they actually visited) and to facilitate comparison between between users we define and use the following metrics based on the desires Long Night visitors reported in interviews as described in [15]:

- **Total event visiting time** – Visitors reported that they want to maximise time spent at events, therefore we summed up the time spent at each visit to get the total visiting time.
- **Ratio of visiting time** – People want to spend as little time getting between events as possible. This metric is the ratio of time actually spent at events over the total duration of the evening.
- **Average event visiting time** – Some users visit only a few events but stay longer whereas others stay for less time to visit more events. The mean of the event visiting time per user will reflect this.
- **Recall of rated events** – If users rate events for possible tour inclusion within the app then a good indicator for success is what ratio of chosen events were actually visited.
- **Precision of rated events** – The ratio of visited events that were chosen by the user beforehand. This indicates how many of the visited events were chosen or discovered through use of the app.
- **Diversity of events** – Visitors reported that they liked the “diverse program on offer,” we calculate the Simpson diversity [19] of the visited events based on the genres they belong to in order to reflect how diverse the event set is:

\[ D = 1 - \sum_{i=1}^{G} \frac{n_i \cdot (n_i - 1)}{n \cdot (n - 1)} \in [0; 1] \]

where \( G \) is the set of all genres, \( n_i \) denotes how often genre \( i \) is present in the set of events and \( n \) being the sum of all \( n_i \). A higher value means greater diversity.
- **Temporal contiguity of events** – The temporal contiguity of a set of events is defined in [15] as “the time needed for a route utilising all available transportation links and visiting all events of that set”. Opening hours and other constrains are not considered. Normalised by dividing by the number of events minus 1.
- **Ratio of top N events** – By counting how often users rated or selected an event it is possible to estimate its popularity. By ordering the events by their popularity a popularity rank can be obtained. A metric for calculating the popularity of the visited events is the number of visited events which are among the top \( N \) events. By normalising this value by the number of visited events we get the ratio of top \( N \) events in the visitor’s itinerary.

6.2 Identifying visitors’ itineraries

A prerequisite to calculating the metrics in the previous section is the ability to extract, for each user, a sequence of event visits from the logged data. In[18], a statistical method for automatically establishing which places were visited was proposed. The model makes use of several features including the mobile phone location information and accuracy, event position data, event opening hours, the popularity and rating of the event over all users, as well as user interactions with the event. We evaluated the approach with our interaction data using gold standard values for visited events based on human labellers. We found the best model showed very accurate performance. For the 111 users who

---

2 On the Android-Platform the user can globally disable GPS

3 Upon completion of the survey they were rewarded with a chance to win a 15 Euro Amazon voucher.
7. GENERAL USER BEHAVIOUR

In this section we look into how users interacted with the app and how those with GPS data spent their evening.

7.1 Behaviour within the app

As the interactions of all 1159 app users were recorded we were able to extract the times the users interact with the app and establish periods of interaction and no interaction. The total time of interaction gives us some insight into how engaged the user was with the app. It is calculated simply by summing the time periods for which a user was active, discounting times where the system reported no interactions for more than 15 seconds. On average, users interacted with the system for 17.21 minutes (median 10.49) during the night. 68.16% of users interacted with the system for more than 5 minutes; 19.24% for more than 30 minutes.

Users had three ways of interacting with the events listed in the various tabs in the app: Reading the description (click-through), rating it for possible tour inclusion (rating) or selecting it for inclusion in the current tour (selecting). We define a “successful operation” as either rating or selecting the event. In total, the users performed 5,489 successful operations on events. 42.4% of which originated from the recommender tab, more than from any other tab, however investigation via A/B testing on previous nights has shown that the default start-up tab has an large influence on tab popularity. Use is spread out quite evenly over the remaining tabs, with the exception of the map tab which was quite infrequently used to select events (4.9%).

The most dominant tab by usage was the Recommender tab which was the primary tab for 37.2% of users. search was the top choice for 24.5%, followed by the by Genre tab (17.4%), the by Tour tab (15.6%) and finally the map tab (5.3%). Users didn’t use all available tabs for selecting events, but instead focused on fewer tabs: 47.7% used only one tab, 33.5% two tabs, 14.2% three tabs and only 4.7% used more than this. Thus, most users (81.2%) tend to stick to one or two tabs from which they select their events, meaning they have clear preferences for specific tabs. Among the tabs used by a visitor, not all are used equally often. Most users preferred to use one tab for event selections; 82.8% of selected events came from the user’s most popular tab, 13.9% came from the second most popular and only 3.3% came from others.

Figure 2 shows the tab usage over the duration of the evening (coloured plot) and the number of active users at the same time point (black dots). Usage of the app peaks between 9 and 10pm and then tails off. This suggests that the app is used largely for event discovery purposes. Moreover, the graphic demonstrates that different tabs were dominant at various points in time. Before 8pm, the participants spent lots of time on tabs useful for event discovery with the aim of planning in mind e.g. searching, browsing and in particular recommendations, which was the default tab.

After this time point, the patterns of tab usage changes. Less time was spent locating events by means of the search and recommendation features with more focus on geographical options. For example, usage of the "By tour" tab remains constant. So if visitors find themselves on the bus route, they can discover other events that are available to them by jumping on the next bus. The tab with the biggest usage increase during the evening is the overview map tab (vertical stripes), which shows events available in the area surrounding the user’s current location. We interpret these patterns as showing that later in the evening spontaneous discovery became increasingly important to the users. This could be because they had visited all of the options they had previously found or because they started later and had no plans or simply wanted to see what other events were available nearby.

To summarise, the system was popular with users with usage peaking at the event opening time of 8pm. The users had clear preferences for particular tabs and tended to stick to these tabs when selecting events of interest. Finally, it seems as if the way the app was used changed as the evening went on. Tabs supporting the spontaneous discovery of events based on the user’s geographical location were used less early on in the evening when events were marked as interesting, but used even more towards the end of the evening.

In the next section we focus on the relationship between a user’s preferred tab and the properties of events that were actually visited.

7.2 Evening outcomes

How users spent their time on the evening can only be determined for those users who had GPS switched on. For the 111 users with a reliable GPS signal we ran our model with the trained parameters to get their event visits. Table 1 shows the mean and standard deviation for the metrics defined earlier.

Table 1: Behaviour of the 111 users with GPS (Diversity and Temp. cont. is only calculated for users with at least two visited events)

<table>
<thead>
<tr>
<th>Metric</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td># event visits</td>
<td>5.05</td>
<td>2.20</td>
</tr>
<tr>
<td>Evening duration (min)</td>
<td>211.0</td>
<td>81.2</td>
</tr>
<tr>
<td>Total event visiting time (min)</td>
<td>127.7</td>
<td>58.8</td>
</tr>
<tr>
<td>Ratio of visiting time</td>
<td>63.7%</td>
<td>22.4%</td>
</tr>
<tr>
<td>Average event visiting time (min)</td>
<td>27.8</td>
<td>14.9</td>
</tr>
<tr>
<td>Recall of rated events</td>
<td>22.4%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Precision of rated events</td>
<td>41.7%</td>
<td>38.4%</td>
</tr>
<tr>
<td>Diversity of events</td>
<td>0.885</td>
<td>0.15</td>
</tr>
<tr>
<td>Temp. cont. of events (min)</td>
<td>10.2</td>
<td>8.3</td>
</tr>
<tr>
<td>Ratio of top 5 events</td>
<td>13.4%</td>
<td>19.6%</td>
</tr>
<tr>
<td>Ratio of top 10 events</td>
<td>22.5%</td>
<td>24.3%</td>
</tr>
<tr>
<td>Ratio of top 20 events</td>
<td>32.6%</td>
<td>26.9%</td>
</tr>
</tbody>
</table>

On average, users visited a total of 5 events, similar to the 4.37 visits reported in [1] and stay at the Long Night for around 3.5 hours, which was considerably less time than the official opening hours (8pm to 3am). Around 2 hours (63.7% on average) of this time is spent visiting events. Consequently, around 83 minutes is spent on other activities, e.g. travelling by bus, walking or waiting. Based on GPS tracks of the users, it appears that most of this time is spent travelling and walking between events. Minimising this time is crucial as interviewed participants in [15] confirmed.

On average, visitors spent 27min at each event, although there was a high standard deviation of 15 minutes. Most
Figure 2: Number of active users are marked with black dots; how their interaction time is spread between different screens/tabs is shown in color. The vertical line marks the start of the night at 8 pm.

Events could be visited at any time during the evening, but some had fixed start and end times, e.g. dance workshops, often lasting 30 minutes, 45 minutes or 1 hour. Survey participants reported that some events were overcrowded or not to their taste causing them to leave earlier than expected, perhaps explaining the large variation.

Recall was very low meaning users rated a lot more events as interesting than it was possible to visit. The 111 considered users rated an average of 8.75 events, possibly a strategy of marking events of interest before the evening like a dog-ear in the booklet in order to then decide spontaneously on the evening where to go next.

The ratios of top N events have large standard deviations, likely because visitors have different opinions regarding popular events. Some may wish to visit predominantly popular events whereas others may instead favour seeking out less "mainstream" events.

8. CORRELATION BETWEEN APP USAGE AND VISITOR BEHAVIOUR

Similar to [15], we examine different features of the system - in our case users who primarily made use of a particular tab. For each user there is a tab which most successful operation originate from, one the second most successful operations originate from, etc. We use this ranking of the tabs to split our users into two groups with regard to a single tab, \( t \):

One group for users whose most used or second most used tab was and another group consisting of all other users.\(^4\)

We consider all users who rated at least one event \((n = 77)\). Table 2 shows the same metrics as before, as well as interaction time with the app and interaction time before the evening, split over these new user groups. All significance tests are performed with either the Z-test if both sets are normally distributed or the Wilcoxon test if they are not (normality determined by the Shapiro test).

**Recommender tab**

Visitors who used the recommender tab most often visit significantly more events than the others \((p = 0.007)\) - nearly 1.5 events more - and spent significantly more time visiting events \((p = 0.023)\). In general, this group used the available time more efficiently, as indicated by a much greater ratio of visiting to travelling time, likely due to the temporal contiguity constraints in the recommender favouring events near to each other. Collaborative filtering also recommends events close to already selected events and since its recommendation is based other users' rating, it can be seen as a form of "wisdom of the crowd". The influence of the recommender can also be seen in the greater percentage of popular events among visits.

Users of the other tabs may have suffered from information overload as they had to consider location, time and topic at the same time, perhaps resulting in less efficient itineraries. The recommender tab, on the other hand, displays only 15 events at a time. It is also probably a good choice for late-decision-makers, i.e. those that do not plan, as it saves them the cognitive load of browsing through all events, as indicated by shorter interaction times, particularly before the evening began (see recommender tab usage in Figure 2 between 4 and 6pm). This suggests that this tab was used more often by people who wanted to build an itinerary for the evening quickly and spontaneously, without the need to spend much time browsing through the available options.

**Tour tab**

Users of the tour tab also visited significantly more events \((p = 0.0469)\), but the average event visiting time is shorter resulting in a similar total event visiting time. As the average evening duration for these users was longer, the actual ratio of visiting time decreases slightly. The temporal contiguity is shorter for the set of events chosen from this tab.

It is our impression that tour tab users restrict the events that they visit to those from one or two of the special bus tours, one participant even reported a strategy of using only a single bus tour, finding only events close to bus stops on the tour. Their explanation for this strategy was the desire to use their ticket for the Long Night extensively i.e. to get the best value for money. If other tour tab users had similar motivations it would explain our finding that these users tended to have longer evenings and spend more time in total visiting events.

As there are normally several events near to each bus stop,
users with this strategy would need to view and evaluate several events before making their decision. This is one explanation for the fact that tour tab users tended to spend more time interacting with the system than other users.

**Genre tab**

[15] reported that the set of events rated by users preferring the genre tab is less diverse, concluding that “users of this tab have very specific interests they want to focus on”. The diversity of the events that were actually visited is also lower for these users, as can be seen in Table 2. The theory that these users are more fastidious about their event choices is further supported by the significantly greater precision \( p = 0.0498 \), meaning they tend to adhere more rigidly to their original plans during the night. They also don’t seem to care about the popularity of events, favouring more esoteric choices that fit more closely with their specific genres of interest. They seem to be willing to trade in their specificity for a smaller number of visited events and also a lower ratio of visiting time and seem to spend more time on preparation.

**Search tab**

People using the search tab spent less time at the events overall - around half an hour less \( p = 0.018 \) - and also stayed 5 minutes fewer per individual event. Since these users also spent less time at the evening overall the ratio of time spent on events is only slightly reduced. In [15] the search tab is reported to be primarily used for known-item search, which means mainly popular events are selected from this tab. This tendency towards popular events during the selection process does not seem to follow through to the actual evening as usage of the search tab seems to have no correlation with the popularity of visited events.

One explanation for this is that the search tab is used by users who want to “cherry-pick” some events they know about in advance, e.g. from advertisements or word of mouth, suggestions from friends or events the user had visited in previous years. Except for these events these users might do no further planning for the evening, as suggested by the low recall. After visiting the events selected before the evening started they might quickly look into some other - less popular, spontaneously discovered - events but end their evening earlier than others if nothing appropriate could be found. This more extensive early planning to pick out known items is reflected in the significantly increased interaction time in the hours leading up to the events \( p = 0.019 \).

**Map tab**

A final user group of interest - from the analyses in Section 7.1 - were users who made use of the map tab to geographically find events of interest, a behaviour which became more common as the evening continued. We wanted to establish how this strategy related to evening outcomes.

For the following analyses we split the user groups based on interaction time spent on each screen rather than the successful operation metric used above, as we know that the map tab was used for event discovery, but far less often for marking events as interesting for plan generation. We compared users who spent most time on the map tab \( n=31 \) vs all other users \( n=90 \). The results show that the map tab users interacted less before the evening (5.6min vs 15.7min; \( p=0.026 \)), which confirms our suspicion that these users did less planning. The events visited by map users were also less likely to have been previously marked as interesting by the user. However, this is likely explained by such users marking fewer events as interesting in the first place (4.71 events vs. 9.79; \( p=0.01 \)).

Other trends in the data for this group include that the temporal contiguity for visited events is lower (12.4min vs. 14.9min), meaning that the visited events were closer together. This makes sense if they were using relying on geographical discovery. Moreover, the events visited were less popular with the entire user population, which again makes sense (10.1% vs. 15.7% of visited events were among the top 5 popular events).

Figure 3, which shows the precision of events marked as interesting over time, demonstrates how the behaviour of the two groups differs. The map users, who had few events in mind to visit, stuck to these events until roughly 9.30pm before precision began to tail off. Whereas the other user

<table>
<thead>
<tr>
<th>Metric</th>
<th>Recommender tab</th>
<th>Tour tab</th>
<th>Genre tab</th>
<th>Search tab</th>
</tr>
</thead>
<tbody>
<tr>
<td># event visits</td>
<td>Rec. tab ((n=40))</td>
<td>other ((n=37))</td>
<td>Tour tab ((n=19))</td>
<td>other ((n=58))</td>
</tr>
<tr>
<td>Evening duration (min)</td>
<td>224.6</td>
<td>201.2</td>
<td>232.4</td>
<td>207.1</td>
</tr>
<tr>
<td>Total event visiting time (min)</td>
<td>114.6</td>
<td>111.7</td>
<td>135.7</td>
<td>126.5</td>
</tr>
<tr>
<td>Ratio of visiting time</td>
<td>67.2%</td>
<td>60.1%</td>
<td>65.5%</td>
<td>64.8%</td>
</tr>
<tr>
<td>Av. event visiting time (min)</td>
<td>28.57</td>
<td>26.55</td>
<td>21.89</td>
<td>29.47</td>
</tr>
<tr>
<td>Recall of rated events</td>
<td>31.8%</td>
<td>32.8%</td>
<td>37.7%</td>
<td>30.5%</td>
</tr>
<tr>
<td>Precision of rated events</td>
<td>61.5%</td>
<td>58.6%</td>
<td>57.5%</td>
<td>60.9%</td>
</tr>
<tr>
<td>Diversity of events</td>
<td>0.884</td>
<td>0.869</td>
<td>0.931</td>
<td>0.860</td>
</tr>
<tr>
<td>Temp. cont. of events (min)</td>
<td>9.80</td>
<td>9.83</td>
<td>8.71</td>
<td>10.17</td>
</tr>
<tr>
<td>Ratio of top 5 events</td>
<td>14.2%</td>
<td>11.3%</td>
<td>18.2%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Ratio of top 10 events</td>
<td>22.4%</td>
<td>20.6%</td>
<td>25.6%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Ratio of top 20 events</td>
<td>35.7%</td>
<td>24.2%</td>
<td>31.8%</td>
<td>29.7%</td>
</tr>
<tr>
<td>Interaction time (min)</td>
<td>37.9</td>
<td>39.0</td>
<td>47.0</td>
<td>35.0</td>
</tr>
<tr>
<td>Interaction time before LN (min)</td>
<td>17.0</td>
<td>19.4</td>
<td>21.4</td>
<td>17.1</td>
</tr>
</tbody>
</table>
group had comparatively high levels of precision until midnight. This could mean that they finished their plans quicker than expected or were less than impressed with their choices. Nevertheless, both groups had the tendency to diverge from their plans as the night went on. This finding relates to and builds on those of [13] who suggested users deviate from plans, but didn’t say how or why.

To summarise the findings of this section: Users of the recommender tab visit many events, stay longer at events and have an efficient itinerary while users of the tour tab visit even more events, stay for a much shorter time and have a less efficient itinerary. Both stay longer on the evening as a whole and have a longer overall visiting time. In contrast, genre tab users and search tab users both visit fewer events and their total evening visiting time is shorter. Overall it seems that there is some similarity between the visiting behaviour of users who favour the recommender and by tour tabs and between those who choose from the genre and search tabs. We also identified a further group of users who made of use of the map tab and who did little in the way of planning. These users visited less popular events and stayed within a small area of the city.

9. QUESTIONNAIRE DATA

To find out more information about how users interacted with the app we constructed a short online questionnaire which 55 visitors of the Long Night responded to.

Table 3 shows how participants responded when asked what factors are important to them when choosing an itinerary. All of the factors seemed to be important to most users, however ensuring efficient use of their time and choosing interesting events seemed most crucial. Table 3 also shows how many participants agreed with various statements about the app. More than two thirds of users agreed with the statement that the app helped them to achieve this as they report that it assisted them in visiting more events (0.26 vs. -0.25) and travel less (0.37 vs. -0.19). Overall, the assistance provided by the recommender tab seems to have pleased these users. As seen before, tour tab users appear to be similar to those of the recommender tab: They want itineraries with an efficient use of time (0.86 vs. 0.24), shorter paths (1.00 vs. 0.64) and featuring many events (0.43 vs. 0.08). They agree with the statement that the app helped them to travel less (0.63 vs. -0.04), however they do not agree so strongly that the app helped them to visit more events (0.13 vs. 0.00). They also value interesting events less than users of other tabs (1.14 vs. 1.38), aligning with the best value for money strategy suggested by their GPS data.

Users of the genre tab were less interested in using their time efficiently (0.22 vs. 0.43), did not seem to care much about having short travel times (0.56 vs. 0.78) and were not bothered about visiting many events (-0.22 vs. 0.30). Instead, they put value on visiting interesting (1.60 vs. 1.22), but not diverse (0.11 vs. 0.65), events and report that discovering new events was enhanced by the app (0.60 vs. 0.40), probably in those users’ specific field of interest.

Search tab users seem to have fewer requirements in general as they consistently value the stated properties equally or less than the control group. Some of these participants commented that they want to have itineraries that make it possible to reach events in time, however their “cherry-picking” of events can lead to inefficient routes and therefore longer travelling times. This is also reflected in the strong disagreement with the statement that the app helped to reduce travelling time (-0.80 vs. 0.48). The search tab seems to have helped those users in cherry-picking, however not being able to make better choices for the user in terms of temporal contiguity resulted in less time spent at events and more time getting between them.

For the questionnaire we again treated the map tab separately by not using number of successful operations but rather interaction time to separate two survey participants into groups: those who spent most of their interaction time on the map tab (n=9) and others (n=37). When asked about how they used the app, 88.9% of the map tab users used it as an electronic program guide whereas only 62.2% of the others did so. This again reflects map tab users having

![Figure 3: Visited events precision over time: Map tab users’ precision declines earlier](image-url)
no ambitions of making plans but instead to spontaneously decide where to go next.

In summary, the questionnaire data reveal several (although due to the small sample sizes not significant) trends. It seems different types of visitors can be distinguished from the results. Visitors have different priorities regarding what they expect from a good itinerary through the night and put different emphasis on having a efficient route, visiting interesting events and having wider or narrower focus on the type of events visited. Our app offers different kinds of assistance in an attempt to cater for the needs of different types of visitors. The questionnaire results show that most users perceived a benefit from the assistance offered, however the perceived benefit correlates with how they chose to use the app on the night.

10. DISCUSSION

We now try to make sense of our findings by comparing the log analysis and questionnaire findings across tabs. We discuss what the findings mean with respect to those reported in previous work and explain why they are useful for designers of assistance systems for distributed events.

Previous work [15] showed a correlation between tab usage and selected events. This, however, does not necessarily evidence assistance being provided to the user. It may have been that the selected events were not actually visited or that the users were not entirely satisfied with the events they selected. Our results in this paper go beyond these in that we move from in-app behaviour to what the users did during the distributed event, showing a correlation between tab usage and events the user actually visited. The results demonstrate that different system usage correlates to different outcomes as reflected in the metrics.

Similar to results in the HIB literature, our findings indicate that people are able to adapt their behaviour. Whereas previous studies have shown that people adjust their behaviour based on the strengths and limitations of the system; when the search algorithm is poorer, they type longer queries or check more results. Here, the trends in the questionnaire data suggest that users are able to adapt their strategy by choosing the best tool for their particular needs. Our findings endorse the recommendations of [5], who promote flexibility in mobile interface design as a single interaction paradigm may not optimally suit all users’ needs.

Moreover, we revealed a second type of adaptive behaviour. Later in the evening, when less time is available for travel or planning, the users start to make use of different features that require less planning and rely more on geographical discovery. A second group of users (map users), who were unable or unwilling to invest time in planning for the festival, used this strategy from earlier on in the evening. In the same way as the filers vs pilers reported in [14], who invest earlier in the hope that later costs will be reduced. Therefore, choice of strategy (planning vs not planning) reflects a trade-off between effort up-front and benefit later on.

When examining the behaviour of the different user groups, a further trade-off can be seen between personal, topical interest and efficiency in itineraries. Users who favoured the recommender and tour tabs seem to be more open to assistance or outside influence when making plans, whereas people who favoured the search and genre tabs seemed to have a better idea of what they wanted and wish to maintain more control over their evening. In terms of the outcomes they achieved, it seems that both the recommender and tour users were able to visit more events and spend more time at events than other users, suggesting that being less flexible towards the genre of events results in a less efficient plan. It is possible, considering both of these highlighted trade-offs, that combining the strengths of different features might be advantageous. This could be achieved, for example, by highlighting (or ranking) events on the map tab based on the user’s recommendation profile or allowing users in the other tabs to sort lists by either distance or estimated personal relevance.

We believe that while our app deals with a specific usage context, our approach here has relevance to the evaluation generally. Going beyond click-through data to look at which events were actually visited was important in our scenario because, as we have shown, many more events were marked as interesting than were actually visited, particularly for non-map users. It provided - and we are unaware of any other work in the literature to achieve this - insight into the relationships between intention, behaviour and outcomes. One can think of several examples in both search and recommender system contexts where our approach would be beneficial, both location detection, e.g. as geographical queries in mobile search or restaurant recommendations, as well as more generally, e.g were any of the found recipes actually cooked, eaten and enjoyed? Or did the exploratory search to find out more about a political party change the way the searcher voted?

11. SUMMARY AND FUTURE WORK

In this paper we performed and described a large-scale user study investigating how information behaviour related to users’ experiences of a distributed event. To help analyse the data we derived a semi-automated method of identifying the events visited by individual users based on log data and
GPS tracks. By analysing user behaviour and, crucially, determining the true outcomes of the evening (i.e. the events they actually visited), we identified the existence of different types of visitors (planners vs. non-planers) who exhibited different needs at different times. Furthermore, we observed that over the course of the event users often adapted their behaviour as information needs became more immediate and localised. The evidence suggests that our app provided different means for these user groups to achieve what they wanted and seems to be helpful. Moreover, complementing previous work, we identified new types of adaptive behaviour: adapting tool choice to suit needs and adapting behaviour over time.

To sum up, the information behaviours of users and how their behaviours relate to their experiences of distributed events like the Long Night of Music are only partly understood. This work has attempted to uncover how their interactions with the information systems provided influenced their evening. We showed that different variables influenced the outcome of their evening: Usage of the system, time and various user types (planers vs. non-planers, interest vs. efficiency). However other potential influences remain to be investigated: Under which circumstances do visitors deviate from their original plan? Are the same criteria important to them when choosing alternative events as for the plans made beforehand? Is visiting alternative events in the middle of the plan different from visiting further events after the plan was completed?

With a better understanding of how visitors behave, we hope that further adaptions of the system to provide better assistance will be possible. In the near future we plan to expand our survey on which route properties visitors consider important. With more data, we hope to be able to test if the trends are significant. Finally, we intend to examine, in more detail, the trade-off between planning and not planning as well as the trade-off between visiting interesting events and events being well located.

Acknowledgments
This work was supported by the Embedded Systems Initiative. We also like to thank Leif Azzopardi from the University of Glasgow for his thoughtful and constructive ideas.

12. REFERENCES