Running head: EVENT PERCEPTION IN MOBILE INTERACTION

Event Perception in Mobile Interaction: Toward Better Navigation History Design on Mobile Devices

Yanqing Cui

Nokia Research Center, Helsinki, Finland

Antti Oulasvirta and Lingyi Ma

Helsinki Institute for Information Technology (HIIT), Helsinki, Finland

Abstract

In this paper, we explore how people perceive interactive activities in order to inform navigation history design on mobile devices. Following event segmentation method, we asked 12 participants to break six episodes of mobile interaction into segments, organize the segments, and identify those deemed representative. Three findings emerged. Firstly, when making sense of mobile interaction, users concentrate on the content objects on which actions are performed. This indicates the value of content-centric designs in navigation history and other mobile UI designs. The content objects are data objects and their collections meaningful to the person dealing with it, for example, photos, messages, or albums. Secondly, users tend to employ two-level hierarchies in grouping segments, and use the similarity in content objects and applications as a reference. They deem the segments as representative where objects are created or changed, or where sharing or querying acts take place. These findings indicate how a navigation history design should organize and prioritize mobile interaction events. Finally, event perception shows relatively low inter-participant consensus, which indicates that navigation history designs have to accommodate large individual differences.

Event Perception in Mobile Interaction: Toward Better Navigation History Design on Mobile Devices

Imagine seeing somebody using a mobile Web browser intensively to find information about, say, flight departures from the Boston airport. The episode would probably take somewhere between 10 seconds and 10 minutes to complete, during which you would see the user clicking links, typing information into search fields, clicking buttons, scrolling up and down, etc. Now, another observer could perceive exactly the same events in some other way, perhaps in terms of navigation, querying, browsing, or searching for information.

Understanding how people perceive interactive activities could be useful input for navigation history design. A design of navigation history typically collects, organizes, and presents past views or a summary of past views for users to continue or repeat their tasks (e.g. Heer, Mackinlay, Stolte, & Agrawala, 2008; Kaasten, Greenberg, & Edwards, 2001; Tang, Lin, Pierce, Whittaker, & Drews, 2007; Tauscher & Greenberg, 1997; Won, Jin, & Hong, 2009), or to navigate in a user interface (UI) (e.g. Park & Kim, 2000). If a design follows the principle of how people perceive their activities, the design has better chance to be well understood and used. *Event perception*, the psychological branch exploring these cognitive principles, however, rarely addresses interactive activities. Find a review on event perception in (Zacks, 2008; Zacks & Tversky, 2001). This study aims to fill the gap. It was conducted as input to designed history mechanisms in the Linked Internet UI concept (Cui, Honkala, Pihkala, Kinnunen, & Grassel, 2010).

In this study, we explore event perception in mobile interactions, focusing on three key research questions. *Question 1:* What are the natural units of interaction for users? For example, the Web page or its counterparts might be a natural unit in a navigation history design, but we do not know whether users perceive events on the basis of interface elements or something else. *Question 2*: How do users prioritize and group elements of a perceived episode of mobile interaction? These cognitive rules reveal the key events and their structures that a navigation history design needs to focus on. *Question 3*: To what extent do users agree with each other in their perceptions? A universally usable interaction demands reasonable *consensus*, or all users mostly segment and organize perception of a given mobile event in similar ways.

We used *event segmentation*, a common method in event perception, in this study. The method usually involves asking observers to segment films of events, such as a person making a bed or assembling a saxophone (Newtson, 1973). Applying it to our context, we asked 12 participants to annotate videos of someone using a mobile device to complete everyday tasks. They had to identify units in the videos, organize the units into structures, identify the units deemed to represent each video well, and verbally explain their logic.

The contribution of this paper is threefold. To our knowledge, the study is one of the first empirical studies that explore principles of how users understand mobile interactions. It is driven by a design purpose, and its findings are interpreted in the context of navigation history design. Finally, the study introduces several methods to calculate the extent to which the participants agree with each other, which can apply to all types of event perception studies.

The paper is organized as follows: We start with a literature review highlighting some key research on event perception. After that, we explain our research method and the result of our study. Finally, we discuss the research and design implications in the end of the paper.

Related Work

Event Perception

Event perception researchers explore "set of cognitive mechanisms by which observers pick out meaningful spatiotemporal wholes from the stream of experience, recognize them, and identify their characteristics" (Zacks, 2008). One common method employed is to ask observers to segment footage of events into units as they watch them, or the event segmentation technique. The breakpoints – reported by observers when they perceived one unit ending and another beginning – are analyzed to reveal event perception principles (Newtson, 1973).

A recurring finding has been that people naturally perceive goal-oriented activities as consisting of discrete parts and subparts (Hard, Lozano, & Tversky, 2006; Kurby & Zacks, 2008; Zacks, Tversky, & Iyer, 2001). For example, "making a bed" is a continuous activity but may be perceived as consisting of "putting on the sheet" and "putting the pillows in their cases." "Putting on the sheet," as a part, has its subparts, such as "unfolding the sheet," "spreading it out," and "straightening it out." There is evidence that observers parse an activity into parts on the basis of distinctive sensory factors in a bottom-up process (Zacks, 2004) and proceed from their understanding of actors' goals in a top-down process (Hard et al., 2006). Most likely, the sensory factors and conceptual interpretations are correlated: when one goal is completed and pursuit of another initiated, there are also physical changes, and vice versa (Kurby & Zacks, 2008).

Some studies suggest a hierarchical structure of event perception: *coarse-level*, using the largest units, and *fine-level*, using the smallest natural units (Hard et al., 2006; Zacks et al., 2001), with, occasionally, a *medium-level*, using the granularity in between (Iqbal & Bailey, 2007). Coarse-grained structural breakpoints typically represent the introduction of objects and broad actions applied to them, while the breakpoints in fine-grained perception are more precise actions applied to the same objects. The two kinds of breakpoints coincide at the same locations significantly more often than could occur by chance alone, which indicates that event perception indeed follows a hierarchical structure (Zacks et al., 2001). According to *common coding* theory, the same schemata underlie the perception of others' actions and the planning of one's own actions (Hommel, Muesseler, Aschersleben, & Prinz, 2001). This implies that people will apply a similar hierarchical structure whether they are perceiving the event online, recalling it after watching it, or even imagining the event from telegraphic descriptions (Zacks et al., 2001).

Interactive Activities

The findings from event perception studies have been applied to the field of human–computer interaction (HCI). The finding that an event is perceived as a structure of discrete units has been used to explain why animations have not yielded better performance than their equivalent static diagrams have in various contexts (Tversky, Morrison, & Betrancourt, 2002). The finding that a breakpoint reflects internal transitions in perception or cognition between two meaningful task units has been used to infer an opportune time for presenting interrupting notifications (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2007). The finding that a breakpoint is typically characterized by a detectable change in sensory stimulus has been used to enhance computer vision work (Rui & Anandan, 2000).

In contrast with the popularity of research on *applying* event perception principles, there are few empirical studies on *exploring* the principles of interactive activities themselves. In one exceptional case, Adamczyk and Bailey (2004) used the event segmentation technique to elicit task models in order to infer optimal interruption times. They asked 25 participants to segment three videos (of document editing, video watching, and Web searching) through breakpoints at coarse and fine levels. The researchers confirmed that observers can break an interaction sequence into a hierarchical structure, and they suggested weighing the potential interruption cost at each breakpoint by measuring how much the users agree with each other. A follow-up study was done by Iqbal and Bailey (2007), who asked 24 participants to segment three videos (of document editing, image manipulation, and programming) at coarse, medium, and fine level. The results showed that coarse- and medium-level breakpoints are associated with switching between applications; coarse-level breakpoints indicate switches in unrelated applications, whereas medium-level breakpoints indicate switches to related applications. Fine-level breakpoints were tied to actions applied to the content within an application – for example, completing formatting commands, searches, and copy-and-paste sequences in a document editing video. Observers did not identify events with finer granularity, such as completing a specific sentence or line of code, or moving between fields in a dialog.

In addition to the event segmentation, protocol analysis is also used to study how people mentally construct interactive activities (Gray, 1990). The difference is that protocol analysis focuses on the actor's instead of the observer's perspective. Byrne, John, Wehrle, and Crow (1999) analyzed verbal protocols from eight participants when they used the Web, and further proposed a task taxonomy that categorized their interactive activities at a granular level. The categories were: "react to environment," "configure browser," "provide information," "go to page," "locate on page," and "use information." The taxonomy categorizes user activities at a granularity level that fits well with our goal in designing a navigation history. Taylor et al. (2008) analyzed voice-mail messages of mobile Internet access sessions and grouped all such sessions into categories: information seeking, information exchange, and action support. The study used *individual sessions* as the units of analysis, which was too abstract for designing a navigation history. The same session-based analysis was found in other relevant studies (Church & Smyth, 2009; Cui & Roto, 2008; Sohn, Li, Griswold, & Hollan, 2008).

Our Approach

In this study, we used event segmentation as the research method after some changes. Firstly, our study used both event segments and breakpoints as units of analysis, thus departing from the approach of earlier studies, wherein the analysis focused on only breakpoints. After identifying points where one unit perceivably ended and another began, the users had to recognize, label, and organize segments between neighboring breakpoints. Secondly, we conduct a mixture of qualitative and quantitative analysis to ensure high ecological validity of this study. The observers had to articulate their rationale after annotating the videos. Thirdly, we experimented with several methods to gauge between- observer consensus when analyzing quantitative data. The metric reveals the extent to which people agree with each other in their perceptions.

We decided to follow the event segmentation approach after considering other alternatives, one of which was task analysis. The point of task analysis is to identify structure in activities where human performance is critical. Typical elements of the analysis include the hierarchy (breakdown of tasks into subtasks) and sequence (which tasks follow each other) necessary for reaching a goal. While there has been discussion of the definition of a task, most task analysis methods do not consider the user's perception to be the defining factor (e.g., Diaper, 2003); instead, they look at deviations from known good solution paths. In this paper, we examine a situation wherein there is no explicit goal but the observers must make sense of actions taken by others. The second alternative approach was observation and interviews. We used observation and interviews in a pilot trial with six participants, asking them to tell their view of tasks that they had recently completed, and the tasks we asked them to perform. They tended to describe interactive activities at too general a level when relying on free recall, or at too fine-grained a level when given a specific task to perform. Some early studies also revealed that it is challenging for users to make sense of an adaptive design while using it. Goren-Bar, Graziola, Pianesi, and Zancanaro (2006) argued that “the fact of being involved in direct interaction with the system often makes it difficult for the subjects to perceive and assess certain aspects that are of a more holistic nature.”

Method

Participants

A dozen participants (six male, six female) were recruited through a mailing list and an online forum from Helsinki University of Technology. The participants were 24 to 36 years old (M = 30.4, SD = 4.0). All had undergraduate education or higher. All used the relevant applications and services shown in the test videos at least several times a week, and 10 of them owned and used a Smartphone at the time of our study. Each participant spent about 90 minutes in the study and received a 20-euro gift card as a reward.

Materials

Six test videos plus one practice video were used in the study. Each test video lasted about 5 minutes and focused on typical mobile activities: information seeking, communicating, and personal space extension (i.e., experiencing personal content) (Cui & Roto, 2008). We chose these task types on the basis of previous studies of how frequent various activities are on mobile devices. The tasks reflected four of the five activities people perform most often on mobile devices: messaging, multimedia operations, personal information management (including use of contact, call log, and calendar applications), and browsing (Verkasalo, 2009). The final task type in the top five, voice calls, was not included because it involves limited user interaction and is difficult to videotape. By contrast, we intentionally emphasized Web-relevant activities to reflect their rapid growth trend (Cui & Roto, 2008; Verkasalo, 2009). The mobile Web-relevant activities are to access the Web through a mobile Web browser or dedicated Web applications.

Table 1 lists the basics of each video. The videos were created by strictly following a set process: We developed scenarios and scripts on the basis of previous studies, agreed on the details within the research team, and finally shot the video through a miniature camera attached to a mobile phone. All videos were used in the study without editing.

PLEASE INSERT TABLE 1 HERE

When preparing the video materials, we considered but did not use videos of users' actions in their natural environment. This choice was made because videos prepared to an agreed script have been used and proven valid in previous event perception studies (Zacks et al., 2001). In addition, it is technically difficult to capture videos of mobile interactions without salient influence. It appears feasible to obtain material on PC activities at an agreed time and place (Czerwinski & Horvitz, 2002); however, the same process would not produce quality videos of mobile interactions.

Procedure

The user study was done in a usability lab, where a participant watched and annotated videos displayed on a computer monitor, then sorted the printed labels on a table. Each participant annotated one practice video and a pair of test videos: V1 (Symbian) and V4 (Facebook), V2 (map) and V3 (Gmail), or V5 (Google) and V6 (Flickr). Each pair of test videos was annotated by four participants throughout the study. Within each pair, the order of the two videos was balanced between participants. The study followed the above design primarily on account of time constraints. The initial plan was to have each participant go through all six videos, but the pilot test showed that a participant can get through only two videos, or one 10-minute video, in a single 90-minute session.

PLEASE INSERT FIGURE 1 HERE

Each participant had to perform three steps to annotate each video.

Firstly, the participant watched the video and reported breakpoints, described by the experimenters as “where one basic meaningful and natural unit ends and another begins.” To report a breakpoint, the participant pressed a button to pause the video and described the basic unit orally. Each description was transcribed by a moderator immediately. See Figure 1(left) for a screenshot of the video annotation tool.

Secondly, the description of each unit was printed on slips of paper, and given to the participant to “build a structure and present what is seen in the video.” The metaphor of a book’s table of contents was used to explain this task. The participant also had to verbally explain the organization logic afterwards. Figure 1(right) shows a participant organizing printed labels.

Finally, the participant identified the units that well represented the video. The instruction was to mark any video element “that deserves to be logged, supposing that the device can record all occurrences, and given a shortcut for a user to see or do it again.” The video elements marked are referred to as *representative* events in this paper.

PLEASE INSERT FIGURE 2 HERE

Figure 2 illustrates the result from one annotation session. The participant first divided a video into seven segments (s1–s7), further organized these segments into a two-level hierarchy with three coarse-level events (c1–c3), and finally identified two segments as representing the entire video well (s1, s5). The data gathered from the first step reveal how users identify and characterize basic units from a mobile interaction episode, addressing research question 1. The data gathered from the second and third steps reveal how the users understand a series of mobile interactions when “grouping” or “prioritizing” event segments, addressing research question 2. The data from all steps were analyzed to reveal the extent to which the users agree in their perception, and thus answering research question 3.

Analysis

The study gathers quantitative breakpoint data and qualitative data on event labels and rationale. The qualitative data set was examined through content analysis to reveal the rules behind the patterns identified. The qualitative data were analyzed by means of statistics; in particular, we adopt Cohen’s Kappa index (Cohen, 1960) with some additions to measure inter-observer consensus on basic segments, event groups, and event representativeness.

Kappa index is a descriptive statistic that summarizes agreement as the percentage of the cases on which the coders agree across a number of objects against bare chance. This value is given by:

$$\kappa = \frac{\sum P_{ii} - \sum P_{i.} P_{.i}}{1 - \sum P_{i.} P_{.i}}, \tag{1}$$

where $\sum P_{ii}$ is the observed proportion of agreement and $\sum P_{i.}P_{.i}$ is the chance proportion of agreement. According to Landis and Koch's widely used heuristics (1977), an index of 0.80 to 1.00 is considered "very good agreement," 0.60 to 0.80 considered "good agreement," 0.40 to 0.60 considered "moderate agreement," 0.20 to 0.40 considered "fair agreement," and 0.20 or less considered "poor agreement."

The basic assumptions for Kappa index are that two observers independently classify each of the N independent objects into one of several independent categories. We introduce two methods for converting our data for this analysis. An e*qual-bin method* is used to convert the discrete data such as breakpoints for consensus analysis. We divide each video into equal-sized bins and calculate the Kappa index according to how much observers agree on each bin, for example, in term of the number of breakpoints. An *unequal-bin* method is used to convert the continuous data of event segments for consensus analysis. This is an extension to the equal-bin method. After dividing a video into equal-sized bins, we further divide all bins that contain multiple segments or parts of multiple segments at the breakpoints until each bin contains one segment or part of a segment only. As a result, the Kappa index can be calculated on the basis of whether observers agree on each bin.

Results

Segmenting Mobile Interaction Events

As their first task, the participants had to break each video into fundamental meaningful units, which we refer to as basic segments, or just segments. By examining the result – with statistical analysis of breakpoints and content analysis of user descriptions – we aim to explore how people identify and characterize basic segments from a mobile interaction episode.

Segment durations and inter-observer consensus. In total, the participants reported 406 basic event segments. An event segment lasts 5–107 seconds, with the average length being 17.4 seconds (*SD* = 15.7). Figure 3 presents the distribution of event segment durations for each video. In general, the segment durations show a positive skew distribution. For all videos other than V4 (Facebook), the distributions peak at around 10 seconds, with a small number of segments longer than 30 seconds. The data from V4 (Facebook) show a different pattern from others. The duration of the segments was longer on average, and the distribution covered a broader scope, with less rapid fall-off. This may have been because Facebook supports relatively long Web pages and rich within-page interactions. This finding will be further analyzed in the following sections.

PLEASE INSERT FIGURE 3 HERE

To measure inter-observer consensus, we use Cohen's Kappa index after converting breakpoint data through the equal-bin method. For a given bin, observers' consensus is 0% if one observer reports breakpoints but the other does not; the consensus increases if both observers report time boundaries but with different numbers; and the consensus is 100% if both observers have reported the same number of breakpoints. These scores were organized into a contingency table: along one diagonal, the cells show agreement between two observers, while the cells along the other diagonal represent disagreement. *Kappa values* are then derived by computing the observed agreement and the probability of chance agreement on the basis of the contingency table. Bin size has an impact on Kappa value. Instead of using a fixed size, we explore Kappa values' distribution curve against bin sizes. As shown in Figure 4, the Kappa index peaks at 0.35, indicating “fair agreement” between observers, when the bin size is 22 seconds. The low Kappa value was caused by the fact that some participants appeared to segment the video with finer granularity than others. For example, one participant broke V3 (Gmail) into 34 segments while the others broke it into 15 or fewer segments.

PLEASE INSERT FIGURE 4 HERE

Basic segments and page breaks. Web pages are the basic units of the Web. Extending the metaphor to non-Web designs, we introduce the “*page*” concept. A page refers to “the largest gestalts (rectangles) that an application shows” (Lauesen, 2005, p. 246), which include Web pages as one example. The page appears to be a logical unit in interactive activities. We were interested in whether observers’ segments tend to co-occur with page changes. In order to measure the extent to which people used switching page as a reference in segmenting mobile interactions, we calculate the time distance between each page-switch time point and the nearest segmentation breakpoint. Figure 5 shows the distance distribution for each video. The order of the videos by increasing mean distance is V5 (Google), V1 (Symbian), V2 (map), V6 (Flickr), V3 (Gmail), V4 (Facebook). Overall, the page switches appear to be aligned well with segmentation breakpoints for most videos. The best match was V5 (Google), with an average distance of 1.23 seconds. The alignment is lower for other cases, particularly V4 (Facebook) with its average distance of 13.24 seconds. One possible explanation is the rich within-page interactions of Facebook. For example, “publishing a status update” was perceived as a separate event by some participants; but such an activity does not involve a page switch.

PLEASE INSERT FIGURE 5 HERE

The distance was particularly large for three observers. For example, one of them generated a distance of 8.88 seconds for V5 (Google), in comparison with 1.59 seconds or less for all other participants, and 15.35 seconds in V6 (Flickr), in comparison with 6.71 seconds or less for all others. The data indicate that some participants may use other references than page switch to segment mobile interactions.

Labels of mobile interaction segments. After breaking a video into segments, the participants needed to label them. The vast majority of the labels were verb–noun phrases (92%, 375 out of 406) – for example, in the form “send a message” or “go to the browser.” The remaining labels contained a verb or noun only, but missing phrasal elements were typically self-evident from the context – for example, “go back” for “go back to Google” and “another message” for “read anther message.”

Most nouns on the labels refer to *content objects*, as defined by Lehikoinen, Aaltonen, Huuskonen, and Salminen (2007, p. 53): “content is data that is targeted at human access, including individual data objects and combinations and collections thereof. It is meaningful to the person dealing with it.” The content objects are always accessible to a given user but are not always possessed or controlled by him or her. They are following kinds of content objects. *Personal content* is possessed by the user and accessible to the user only; this includes personal notes and photos. *Social content* is possessed, or was originally possessed, by either the user or other individuals known to the user and is made accessible to the others, examples being Facebook updates and mail messages. *Public content* is possessed by institutions such as news agencies, or by strangers such as Web bloggers.

The other nouns on the labels refer to *UI elements*. They are the *application* and its counterparts that form the structure of the user interface or *UI widgets* that populate the user interface, such as buttons, form fields, and menu items. These UI elements do not have persistent meaning for users other than merely enabling momentary interaction with the devices.

Content objects were far more common than UI elements among the labels given by participants. They were included for 73% of the event segments (298 out of 406), while UI elements were involved in only the remaining 27% (108 out of 406) of cases. Content objects also appeared to be the focus of event perception and therefore were more important than UI elements. As one participant put it, “Well, I only care about the stuff I want to do, not the things I have to do.” It is clear that interacting with content objects was what she wanted to do, and interaction with UI elements was mostly what she had to do.

Mobile interaction taxonomy. The labels given by participants were further analyzed to build a mobile interaction taxonomy. We firstly categorized all labels according to the nouns they contained, into two large clusters: segments involving content objects and segments involving UI elements. For the content-object-related segments, our first categorization was based on the type of the verbs involved, primarily in terms of the effect these verbs have on the relevant objects. This process led to the categories “share content with others,” “create or change content,” “query for content,” and a disproportionately large collection of segments involving verbs that led to no intentional change to the objects. To balance the taxonomy, we further processed the collection. The elements were categorized into “follow social content,” “explore public content,” and “access personal content” according to the type of content objects involved. For UI-element-related segments, we considered the type of UI elements and divided the relevant events into the categories “switch applications” and “finish interaction components.” These categories were of a reasonable size, so we did not divide them any further. See Figure 6 for an illustration.

PLEASE INSERT FIGURE 6 HERE

"Share content with others" refers to the event segment of sending or publishing content for one or more other contacts, such as sending a message or sharing a photo. The event is always intended to reach other people besides the originating user.

"Create or change content" refers to the event segment of creating and modifying content for personal access, or utilizing content to customize devices – for example, adding a contact entry or changing wallpaper. Events in this category typically lead to significant changes to the content but do not involve any other people than the user himself or herself.

"Query for content" refers to the event segment of locating some content through keywords or similar mechanisms. For example, one might search via a web search engine or maps. The event is characterized by the user retrieving a list of search results after entering some keywords or selecting a predefined command such as "find my location" on a map.

"Access personal content" refers to the event segment of viewing, watching, checking, reading, or listening to individually possessed content. Most of the personal content is generated by the user – for example, personal photos or calendar entries.

"Follow social content" refers to the event segment of exploring social content originating with known contacts, such as mail messages and online micro blogs. The users can keep in touch with known contacts by following their content.

"Explore public content" refers to the event segment of checking public content from the Web or other services. Examples are checking a weather forecast and looking at online news. People typically access these for entertainment or information purposes.

"Switch applications" refers to the event segments of entering or exiting an application or Web site merely to set up an environment for other user tasks, such as starting or closing a browser, and back-stepping. We also place device activities that do not involve users in this category, such as a device notifying the user of the battery running low.

"Finish interaction components" refers to the event segments of dealing with UI widgets as one of multiple steps to manipulate a content object. A common case involves editing individual fields in a form UI – for example, "give a title" and "compose detailed description" when one is uploading a photo. These events are quite detailed compared with other categories.

Several measurements were taken to ensure the reliability of the taxonomy. The taxonomy was constructed through consulting prior research, analyzing the capabilities of mobile devices, and assessing the data from our participants. It was then applied to the entire data set by two researchers independently. The inter-coder agreement shows a satisfactory result, a Kappa index of 0.69, or "good agreement."

Grouping and Prioritizing Segments

As their second and third task, the participants had to organize printed labels to build the video structure and to identify the segments representing the video well. The findings answer the second research question on how users summarize a perceived episode of mobile interaction.

Grouping mobile interaction segments. The participants primarily used a two-level hierarchy to structure event segments. The hierarchy contains one *coarse level* above the basic segments (*fine level*). As illustrated in Figure 2, seven basic event segments (s1–s7) are grouped into three coarse-level events (c1–3). In a three-level hierarchy, there is an extra, medium level between the coarse and the fine, but this was seen with only three videos: V2 (map), V3 (Gmail), and V6 (Flickr).

Across all videos, the participants reported 141 coarse-level events and 24 medium-level events. On average, a coarse-level event had 2.9 segments (*SD* = 2.4) and lasted 50.9 seconds (*SD* = 46.7). A medium-level event, existing in only three-level hierarchies, contained 1.5 segments (*SD* = 0.9) and lasted for 33.9 seconds (*SD* = 21.4). Figure 7 shows the duration distribution of mobile events at the three hierarchy levels. Coarse- and medium-level events occur in a broader distribution than do low-level events, or basic segments, as reported above. Their distributions appear to have a similar shape, which could indicate a similar nature.

To measure inter-observer consensus, we use Cohen's Kappa index after converting breakpoint data through the equal-bin method, as explained above. We calculate inter-observer consensus through analysis of breakpoints of coarse-level groups. The Kappa value reaches 0.47, indicating "moderate agreement," with a bin size of 34 seconds. This is higher than for fine-level events. As reported earlier, the inter-observer consensus peaks at 0.35, with a bin size of 22 second, or "fair agreement." The difference indicates the value of summarizing basic segments, since it appears that the users have greater consensus for coarse-level events.

PLEASE INSERT FIGURE 7 HERE

We examined event segmentation results and interview transcripts to explore rationales in grouping of segments. In line with earlier studies, the observers attempted to infer the user's goals in this step: they tended to put together the basic segments that achieve one discrete goal.

In addition to clear relevance to inferred user goals, there were two other salient features that serve as good grouping indicators: similarity in *application* and *content object.*

Event segments were often grouped together when they occurred within the same *application* or subordinate sections thereof. One participant used "*place*" as a metaphor in explaining his logic: "People need to go to a place first, and then do things there. There, things should be put together as a group. When one leaves for the next place, one starts with a new group." "Place" was the primary logic in the grouping of basic segments at a coarse level in three videos. These grouping are Gallery, Inbox, Calendar, Music, and other applications in V1 (Symbian); "Home," "Profile," "Friends," and other sections in V4 (Facebook); and google.com, nokia.com, and other Web sites in V5 (Google). The latter two videos did not involve application switch. Instead, the observers used a subordinate or counterpart of an application as references. A higher-level event typically started and ended with a "switch applications" segment – entering and leaving a "place" – between which were activities to manipulate content objects.

Event segments were also grouped together when involving *content objects* with a relationship recognized as relevant, especially when there was not much change in the application over a longer period of time. Content objects remained of relevance within a segment group but clearly differed from those involved in neighboring segment groups. A segment group may involve identical objects, a set of similar objects, or an object with its subordinates. Overall, content object relevance was the primary logic in grouping of basic segments in three videos: identical message items in V3 (Gmail), points of interest of the same type in V2 (map), and photos in identical albums in V6 (Flickr).

Our analysis highlights the above two grouping logics in our data, but it does not rule out other possibilities. As a matter of fact, an observer may well have perceived user activities as an entangled web and contextually selected one dimension, whether consciously or not, when asked to structure an event. For example, one observer reported facing alternatives in structuring the event. Before structuring a video, he asked, “What do you want me to do – should I relate these to application, content, or what?”

Prioritizing mobile interaction segments. As the last task, the participants had to mark the event segments that represented each video well. “Event representativeness” is operationally defined as the extent to which people perceive an event as worth being recorded in the history, on the assumption that the device is capable of recording all occurrences. In the example shown in Figure 2, the participant marked two of the video’s seven basic event segments as representative. The participants did not perceive events as equally important. Of all basic segments identified, they marked 26% (107 out of 406) as worth mentioning in the history, or as representative under our definition.

Here we use Cohen’s Kappa index to measure inter-observer consensus after we convert segment data through the unequal-bin method. We first divide a video into equal-sized bins and then divide all of the bins further by using breakpoints as bin boundaries where these exist. Each bin is rated as either representative or not, which results in a 2x2 contingency table. The Kappa value of each bin is calculated by computing the observed agreement and the probability of chance agreement. The average consensus for history across all videos and all observers reaches 0.48 at the bin size of one second, which is in the “moderate agreement” range. This indicates that people exhibit some agreement in prioritizing events but the difference between individuals is still large.

As mentioned earlier, we propose an eight-category taxonomy of mobile interaction. A given event’s representativeness is closely correlated with the category to which it belongs;

$x^2(7, N = 406) = 69.15$, $p < .01$. Of the events, 45% or more were perceived as representative in the categories “share content with others,” “create or change content,” and “query for content,” while 6% were perceived as representative in the category “switch applications” and no events were perceived as representative in the category “finish interaction components.” The events in the rest of the categories fell between these percentages. See Table 2 for detailed findings.

PLEASE INSERT TABLE 2 HERE

In the categories “share content with others,” “create or change content,” and “query for content”, 45–62% of the reported segments were perceived as representative. Events in the category of “share content with others” always engage partners, which increased their representativeness. For example, one participant said he sometimes needed a quick access to a sent message in order to justify “I did send you a reminder!” to his girlfriend. Events in the category of “create or change content” always produce results, and the participants claimed a need to monitor the results produced or to reuse them later. Events in the category “query for content” typically involve text entry and the participants generally expected mobile devices to remember all query keywords. As one participant explained, “When I input text, especially queries I’ve made, the task automatically becomes noteworthy. I may search for the relevant stuff again.”

In the categories “access personal content,” “follow social content,” and “explore public content”, 19–37% of the segments reported by participants were perceived as representative. Among these three content types, event segments in the “explore public content” category were rated more representative than those in the other two categories. This appeared to be caused by there being fewer means of access to public content, apart from history, than to other type of content.

In the categories “switch applications”, 6% of the relevant segments were perceived as representative. In the category “finish interaction components”, the percentage was zero. These events were solely to enable interactions, not to manipulate content on their own. As one participant articulated, “The value of these transient steps will vanish after a major action is taken. For example, I may need to go back to earlier steps when creating an MMS. But once I send out the item, I need to check only the sent item, not the steps any more.” Segments such as entering applications seemed to be exceptional cases – for example, the participants mostly marked “log in to Gmail” as representative.

In summary, an event segment was primarily perceived as representative when it involved content objects rather than UI elements. For the content-object-related-events, the categories “share content with others,” “create or change content,” and “query for content” were rated more representative than the others. When asked the reasons behind this, most participants explained that users actively *author* content in these categories, which made them more important than the categories “follow social content,” “explore public content,” and “access personal content,” wherein the users only *read* content, not actively introducing any changes.

Discussion

Interactive activities on mobile devices differ from our everyday physical activities. In our study, mobile interactions primarily rely on virtual interface presentations, whereas physical activities rely on real-life objects. Mobile interactions occur in a strictly designed environment, affording a restricted set of options for proceeding from a given state, while physical activities occur in an open environment with almost unlimited options for any given point. With these differences in mind, we discuss our findings alongside those of previous work. The key findings are listed as follow. Firstly, the participants perceived mobile interaction as consisting of discrete and brief segments. They mostly labeled the segments through verb–object phrases. Most phrases involved content objects; the others only involved UI elements such as application and UI widgets. Secondly, the participants primarily grouped the segments when occurring in the same application or revolving around relevant content objects. They deemed some segments as the most representative wherein an object was created or changed, or where an act of sharing or querying took place. Thirdly, there was a relatively low inter-observer consensus in mobile interaction perception.

Event Perception in Mobile Interaction

Mobile interaction segmentation. In agreement with findings on event perception with physical activities (Zacks et al., 2001), users can break mobile interactions into segments with reasonable consensus. The duration of a segment is comparable with that of everyday activities, but substantially shorter than computer interactions – mobile interaction: 17.4 seconds on average; physical activities: 12.8 seconds (Zacks et al., 2001); computer interaction: 114 seconds (Iqbal & Bailey, 2007). The differences indicate that mobile interaction is perceived at a faster pace than computer interaction is. The page switch appears to be a reasonable reference for breaking up events, but the logic is problematic for some cases, the most prominent one being within-page interactions. Such cases are likely to increase in number with the growth of Web-based applications (Obendorf, Weinreich, Herder, & Mayer, 2007).

In agreement with the earlier studies, content objects play a critical role for people as they segment, structure, and prioritize mobile interactions (Zacks et al., 2001). In mobile interaction, content objects are “data that is targeted at human access, including individual data objects and combinations and collections thereof. It is meaningful to the person dealing with it" (Lehikoinen et al., 2007, p. 53). Among all the labels in describing event segments, three quarters of event segments involved content objects, one quarter of segments were described as involving UI elements only. This salience of UI elements indicates the complexity of current mobile UIs. The UI elements, such as applications, are merely to enable momentary interaction with devices. Therefore, they should not attract much attention from observers in an ideal UI.

Grouping and prioritizing mobile interaction. In line with our initial assumption, a user can group basic event segments into a hierarchy, and prioritize them to report the representative ones. It is of note that there are relatively large individual differences in both cases.

When grouping segments, the participants typically used the similarity of applications and content objects as references. This finding is in line with previous studies on interactive activities. For example, coarse- and middle-level units are found often tied to switching applications (Iqbal & Bailey, 2007). When prioritizing segments, the participants identified about a quarter of mobile interactions as worth mentioning in the history, or representative according to our definition. These are more likely to be the "authoring" segments, where an object is created or changed, or wherein a sending or querying act takes place, than the "reading" segments, where people browse personal, social, or public content. The division between "authoring" and "reading" is of technical origin, but it appears to be a natural response from users. This finding supports the design rationale in utilizing a division such as "edit wear" and "read wear" (Alexander, Cockburn, Fitchett, Gutwin, & Greenberg, 2009; Hill, Hollan, Wroblewski, & McCandless, 1992) and suggests a dimension for future work in exploring, for example, task contexts (see Gyllstrom, 2009).

Design Implications: Toward a Better Navigation History

This event perception study is mainly to inform navigation history and other designs. We are particularly interested in *system-wide* history design that covers all interaction events on a mobile device. This device may follow conventional styles that center on individual applications and their menu hierarchies. It may alternatively structure functions through associations and hypertext navigations, as shown in Linked Internet UI concept (Cui et al., 2010).

Navigation history design. The content object is an essential reference in perceiving mobile interaction. This suggests the value of a *content-centric* design. A design as such needs to monitor the content objects' status in addition to exact page views during the gathering of discrete interaction elements – for example, logging the photo being published as well as the exact steps in publishing it. When optimizing for re-access after a long interval, a design may focus on content objects as users are likely to look for content objects rather than exact views (Cutrell, Robbins, Dumais, & Sarin, 2006; Dumais, Cutrell, Cadiz, Jancke, Sarin, & Robbins, 2003). When aiming for short-term re-access such as back-stepping, a design may still focus on individual views. In this case, the analysis on the content objects' status can aid in suggesting a smaller back-step for meaningful within-page interactions, or, the opposite, devising shortcuts to skip some less meaningful pages (Milic-Frayling et al., 2004). The content-centric rationale can also be used to guide history presentation. An interaction segment is often characterized by objects and user actions applied to them, which indicates that this is a desirable way to present history items as verb–noun phrases such as "send a message" or "play the song named ABC." In the cases of multimedia presentations such as icons, thumbnails, or snippets (Cockburn & Greenberg, 2000; Kaasten & Greenberg, 2001; Teevan et al., 2009; Vartiainen, Roto, &

Kaasalainen, 2008), we propose that the design only needs to visualize content objects if representation of the actions undertaken is difficult and prone to confusion.

There are alternative rationales in organizing and presenting of navigation history. See Figure 8. In addition to the conventional *full list*, a design may *group* relevant views to show them in hierarchical or summary form, or *prioritize* the views to highlight the important ones. These alternative presentations seem natural for mobile user interfaces, in view of the limited display size. After all, it is an overwhelming task to locate a target from a long navigation history even on a conventional computer (Kaasten & Greenberg, 2001; Won et al., 2009).

PLEASE INSERT FIGURE 8 HERE

The study results reveal some principles in grouping or prioritizing interaction histories to shorten history lists. In the grouping rationale, observers summarized basic segments into a hierarchy according to inference of the user's goal, pointing to the relevance of content objects or applications. The findings suggest that machine inference of grouping events can recognize the user's goal as a latent variable, with content objects and applications used as real-time data input. Events can be grouped together when they involve identical objects or a set of similar objects, or when they occur in the same application. In the prioritizing rationale, a navigation history design should consider segment types. A history design may fully cover the content objects involved in the categories "share content with others," "create or change content," and "query for content." It may only highlight the content of landmark values in the categories "follow social content," "explore public content," and "access personal content." See a review of landmarks in the work of Sorrows and Hirtle (1999). For the rest categories, those that involve UI elements only, the design might hide all of them other than those events initiating new applications.

The existence of large individual differences suggests that no history will instantly work for everybody. Thus, a good design has to accommodate the differences. For example, one design may de-emphasize instead of remove items when shortening navigation history. The less important views remain accessible, for display on the press of a “Show all” button. The design should also be easy for all users to understand. As the *apprehension principle* states, the structure and content of the external representation should be readily and accurately perceived and comprehended (Tversky et al., 2002).

Other implications. A mobile UI should concentrate on the content, which is meaningful for users, and minimize the salience of UI elements. This content-centric design rationale is applicable to general mobile UI designs. An ideal mobile UI design should try to simplify the distinctive structures in a device, and make it easier for users to concentrate on the meaningful content objects. For example, the case of Linked Internet UI concept proves the potential of application-less UI on mobile devices (Cui et al., 2010).

The findings on event prioritization indicate the potential of adaptive content presentation. For example, an adaptive design could utilize interaction history to automatically personalize the overall interface presentation for individual users rather than directly present interaction history per se (e.g., Jameson, 2008). This study reveals the most representative events to be those in which an object is created or changed, or in which a sharing or querying act takes place. Accordingly, an adaptive design could prominently present content or actions after being repeatedly involved in representative events, i.e. the acts of creating, changing, sharing, or querying takes place. For example, after frequently appearing in these key events, a contact and its relevant content will be presented more saliently than others.

In summary, we followed the event segmentation paradigm and derived the cognitive rules applied in segmenting, grouping, and prioritizing mobile interaction. This study indicates the general requirement to reduce the salience of UI elements in the future designs, and suggests specific design logics to create a grouping- or prioritization-based navigation history, supposed that the conclusion holds true in the context of interaction design. It also reveals that users do not follow one single principle for event perception, but they may easily learn to utilize it when provided with one design that follows a common principle. This can only be studied by implementing prototypes and letting users work with them long enough in future studies.

References

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: the effects of interruption at different moments within task execution. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, 271–278. doi: 10.1145/985692.985727.

Alexander, J., Cockburn, A., Fitchett, S., Gutwin, C., & Greenberg, S. (2009). Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. *Proceedings of the 27th international Conference on Human Factors in Computing Systems*, *CHI'09*, 1665–1674. doi: 10.1145/1518701.1518957.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled Web we wove: a taskonomy of WWW use. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *CHI '99*, 544–551. doi: 10.1145/302979.303154.

Church, K., & Smyth, B. (2009). Understanding the intent behind mobile information needs. *Proceedings of the 13th International Conference on Intelligent User Interfaces*, *IUI '09*, 247–256. doi: 10.1145/1502650.1502686.

Cockburn, A., & Greenberg, S. (2000). Issues of page representation and organization in Web browser's revisitation tools. *Australian Journal of Information Systems*, *7*(2), 120–127

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46. doi:10.1177/001316446002000104.

Cui, Y., Honkala, M., Pihkala, K., Kinnunen, K., & Grassel, G. (2010). Linked Internet UI: A mobile user interface optimized for social networking. *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI'10*, 45-54. doi:10.1145/1851600.1851611.

Cui, Y., & Roto, V. (2008). How people use the Web on mobile devices. *Proceedings of the 17th international Conference on World Wide Web, WWW '08*, 905–914. doi: 10.1145/1367497.1367619.

Cutrell, E., Robbins, D., Dumais, S., & Sarin, R. (2006). Fast, flexible filtering with phlat. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06,* 261–270. doi: 10.1145/1124772.1124812.

Czerwinski, M., & Horvitz, E. (2002). An investigation of memory for daily computing events. *Proceedings of HCI 2002: Sixteenth British HCI Group Annual Conference*, 229–246.

Diaper, D. (2003). Understanding task analysis for Human-Computer Interaction. In D. Diaper & N. Stanton (Eds.), *The handbook of task analysis for Human-Computer Interaction.* (pp. 5–48). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've seen: a system for personal information retrieval and re-use. *Proceedings of the 26th Annual international ACM SIGIR Conference on Research and Development in information Retrieval*, *SIGIR '03,* 72–79. doi: 10.1145/860435.860451.

Goren-Bar, D., Graziola, I., Pianesi, F., & Zancanaro, M. (2006). The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides. *User Modeling and User-Adapted Interaction 16* (1), 31-62. doi: 10.1007/s11257-006-9004-7.

Gray, S. H. (1990). Using protocol analyses and drawings to study mental model construction during hypertext navigation. *International Journal of Human Computer Interaction, 2* (4), 359-377. doi:10.1080/10447319009525990.

Gyllstrom, K. (2009). Passages through time: chronicling users' information interaction history by recording when and what they read. *Proceedings of the 13th International Conference on Intelligent User Interfaces, IUI '09*, 147–156. doi: 10.1145/1502650.1502673.

Hard, B. M., Lozano, S. C., & Tversky, B. (2006). Hierarchical encoding of behavior: translating perception into action. *Journal of Experimental Psychology: General*, *135* (4), 588–608. doi:10.1037/0096-3445.135.4.588.

Heer, J., Mackinlay, J. D., Stolte, C., & Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1189-1196.

Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '92,* 3–9. doi: 10.1145/142750.142751.

Hommel, B., Muesseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral & Brain Sciences*, *24*(5), 849–937.

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'07,* 697–706. doi: 10.1145/1240624.1240732.

Jameson, A. (2008). Adaptive interfaces & agents. In A. Sears & J. A. Jacko (Eds.), *The Human-Computer Interaction handbook: Fundamentals, evolving technologies and emerging applications* (2nd ed.) (pp. 433–458). Mahwah, NJ: Erlbaum.

Kaasten, S., & Greenberg, S. (2001). Integrating back, history and bookmarks in Web browsers. *CHI '01 Extended Abstracts on Human Factors in Computing Systems, CHI' 01*, 379–380. doi: 10.1145/634067.634291.

Kaasten, S., Greenberg, S., & Edwards, C. (2001). How people recognize previously seen Web pages from titles, URLs and thumbnails. *Proceedings of HCI 2002: Sixteenth British HCI Group Annual Conference*, 247–266.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72–79. doi:10.1016/j.tics.2007.11.004.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 22,* 159–174. doi:10.2307/2529310.

Lauesen, S. (2005). *User interface design: a software engineering perspective*. Boston, MA: Addison–Wesley.

Lehikoinen, J., Aaltonen, A., Huuskonen, P., & Salminen, I. (2007). *Personal content experience: Managing digital life in the mobile age*. Wiley–Interscience.

Marsden, G. & Jones, M. (2001). Ubiquitous computing and cellular handsets – are menus the best way forward? In *Proceedings of SAICSIT'01*, 111-119.

Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A., & Sommerer, R. (2004). Smartback: supporting users in back navigation. *Proceedings of the 13th international Conference on World Wide Web, WWW '04,* 63–71. doi: 10.1145/988672.988682.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28–38. doi:10.1037/h0035584.

Obendorf, H., Weinreich, H., Herder, E., & Mayer, M. (2007). Web page revisitation revisited: implications of a long-term click-stream study of browser usage. *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems, CHI '07, 597–606. doi: 10.1145/1240624.1240719.

Park, J., & Kim, J. (2000). Contextual navigation aids for two World Wide Web systems. *International Journal of Human-Computer Interaction, 12*(2), 193–217. doi:10.1207/S15327590IJHC1202_3.

Rui, Y., & Anandan, P. (2000). Segmenting visual actions based on spatio-temporal motion patterns. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2000*, 111–118. doi: 10.1109/CVPR.2000.855807.

Sohn, T., Li, K. A., Griswold, W. G., & Hollan, J. D. (2008). A diary study of mobile information needs. *Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, *CHI '08*, 433–442. doi: 10.1145/1357054.1357125.

Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In C. Freksa and D. M. Mark (eds.), *Lecture Notes In Computer Science*: Vol. 1661. (pp. 37–50). London: Springer–Verlag. doi:10.1007/3-540-48384-5_3.

Tang, J. C., Lin, J., Pierce, J., Whittaker, S., & Drews, C. (2007). Recent shortcuts: using recent interactions to support shared activities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, 1263-1272. doi: 10.1145/1240624.1240816.

Tauscher, L., & Greenberg, S. (1997). How people revisit Web pages: empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies, 47* (1), 97–137. doi:10.1006/ijhc.1997.0125.

Taylor, C. A., Anicello, O., Somohano, S., Samuels, N., Whitaker, L., & Ramey, J. A. (2008). A framework for understanding mobile internet motivations and behaviors. *CHI '08*

Extended Abstracts on Human Factors in Computing Systems, 2679–2684. doi: 10.1145/1358628.1358744.

Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., André, P., & Hu, C. (2009). Visual snippets: summarizing Web pages for search and revisitation. *Proceedings of the 27th international Conference on Human Factors in Computing Systems, CHI'09,* 2023–2032. doi: 10.1145/1518701.1519008.

Tversky, B., Morrison, J., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human Computer Studies, 57*, 247–262. doi:10.1006/ijhc.2002.1017.

Vartiainen, E., Roto, V., & Kaasalainen, J. (2008). Graphical history list with multi-window support on a mobile Web browser. *Proceedings of the 2008 Third international Conference on internet and Web Applications and Services, ICIW'08*, 121–129. doi: 10.1109/ICIW.2008.13.

Verkasalo, H. (2009). *Handset-based analysis of mobile service usage*. Doctoral dissertation, Helsinki University of Technology, Finland.

Won, S., Jin, J., & Hong, J. I. (2009). Contextual Web history: using visual and contextual cues to improve Web browser history. *Proceedings of the 27th international Conference on Human Factors in Computing Systems, CHI '09*, 1457–1466. doi: 10.1145/1518701.1518922.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*, 979–1008. doi:10.1016/j.cogsci.2004.06.003.

Zacks, J. M. (2008). Event perception. *Scholarpedia, 3* (10), 3837.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127* (1), 3–21. doi:10.1037/0033-2909.127.1.3.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130* (1), 29-58. doi:10.1037/0096-3445.130.1.29.

Author Note

Yanqing Cui, Nokia Research Center, Helsinki, Finland; Antti Oulasvirta, Helsinki Institute for Information Technology (HIIT), Helsinki, Finland; Lingyi Ma, Helsinki Institute for Information Technology (HIIT), Helsinki, Finland.

This study was funded by Nokia Research Center. We owe a special thanks to all of the participants for their patience, to Mika Rautava, Virpi Roto, and Guido Grassel for their help in planning and executing the study, and to Elina Ollila for reviewing one manuscript version of this paper.

Correspondence concerning this article should be addressed to Yanqing Cui, Itämerenkkatu 11-13, Helsinki 00180, Finland. Email: yanqing.cui@nokia.com.

Table 1

The six videos of mobile interaction and the mobile devices where the videos were taken

Video name	Scenario in brief	Device
V1. Symbian	Negotiate an appointment while playing with the mobile phone	Nokia N95
V2. Map	Find a restaurant on a map, and notify others of its address	Nokia N95
V3. Gmail	Read, reply to, and send mail from a mobile Web browser	Nokia N810
V4. Facebook	Make new friends and check on others' status	Nokia N810
V5. Google	Search for facts and opinions about a phone model	Nokia N810
V6. Flickr	Publish a new photo and check new photos from others	Nokia N810

Table 2

Frequency, duration, and representativeness of mobile interaction categories

Category	**Frequency**	**Duration (M/SD)**	**Representativeness**
Segments involving content objects			
Share content with others	54	20.07 / 21.57	50%
Create or change content	20	22.11 / 15.33	45%
Query for content	26	10.06 / 9.19	62%
Access personal content	20	14.65 / 12.61	25%
Follow social content	107	19.88 / 15.64	19%
Explore public content	65	13.84 / 18.00	37%
Segments involving UI elements			
Switch applications	96	18.06 / 15.01	6%
Finish interaction components	18	9.56 / 6.79	0%

Figure Captions

Figure 1. A screenshot from the video annotation tool (left) and a scene when a participant was organizing the printed labels (right).

Figure 2. Illustrated procedure of annotating a video.

Figure 3. The distribution of mobile interaction segments in different videos.

Figure 4. Average users' consensus over all basic event segmentsat varying bin size.

Figure 5. The time distance of a breakpoint from the nearest page switch.

Figure 6. Taxonomy of perceived mobile interactions.

Figure 7. Distribution of mobile interaction segment groups at different hierarchical levels.

Figure 8. The three alternative navigation history designs.

Figure 1.

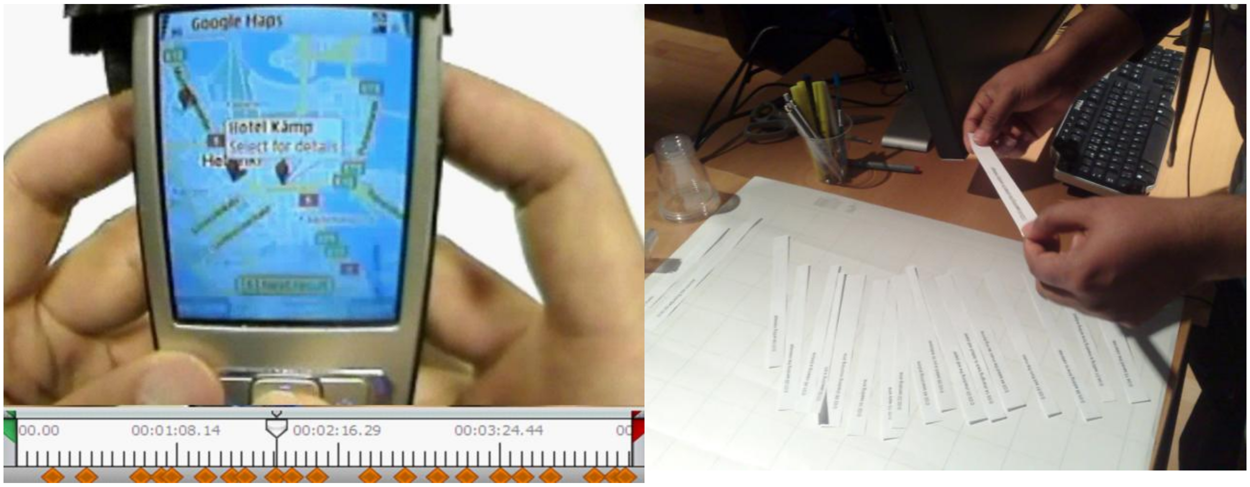
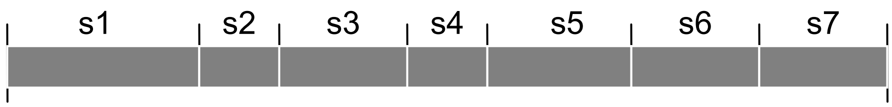
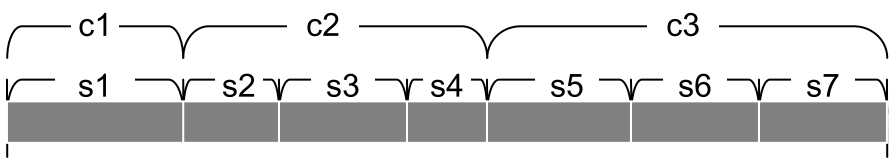

Google Maps
Hotel Kämp
Select for details
00.00
00:01:08.14
00:02:16.29
00:03:24.44

Figure 2.

Step 1: Watch a video to report breakpoint, and orally label each segment (s1-s7). This reveals what the natural units of mobile interaction for users are.



Step 2 and 3: Organize printed labels to build the video structure (up), and identify the segments that represent the video well (below). The highlighted items (s1 and s5) are deemed as representing the event well.



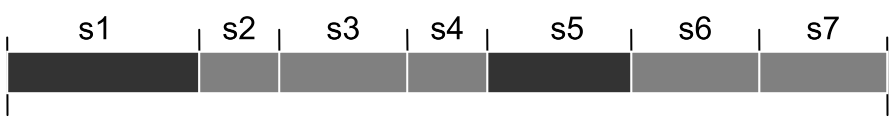


Figure 3.

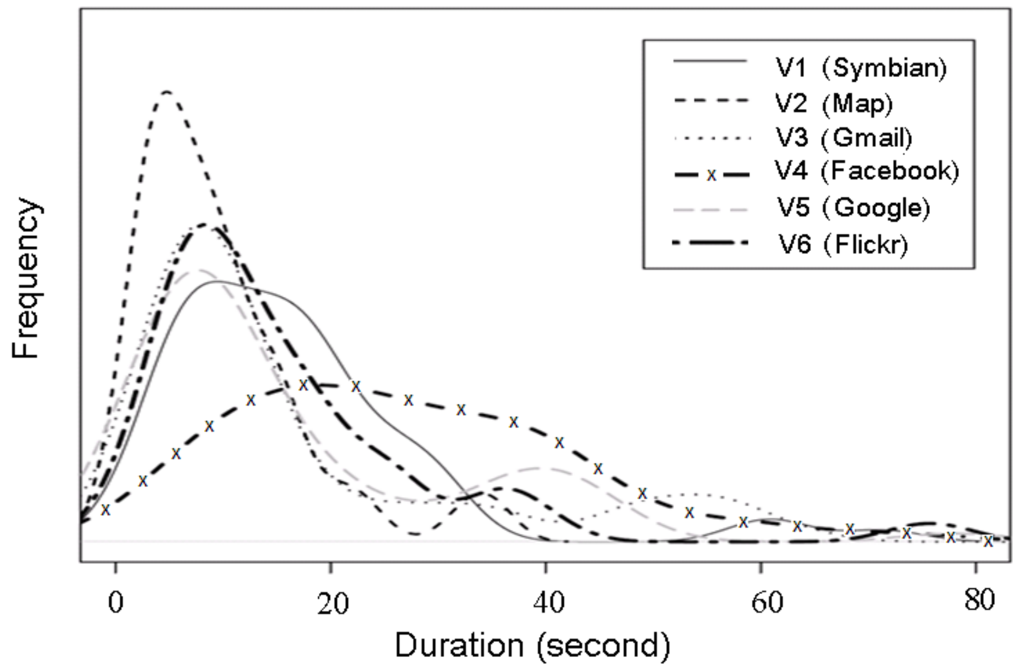

V1 (Symbian)
V2 (Map)
V3 (Gmail)
V4 (Facebook)
V5 (Google)
V6 (Flickr)
Frequency
0
20
40
60
80
Duration (second)

Figure 4.

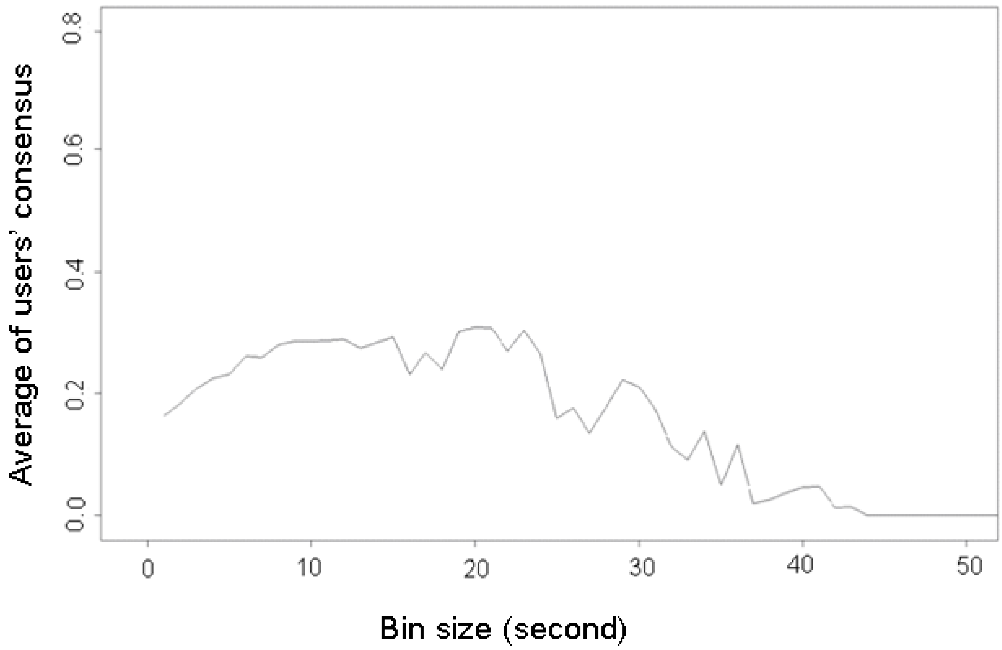

Average of users' consensus
0.0
0.2
0.4
0.6
0.8
0
10
20
30
40
50
Bin size (second)

Figure 5.

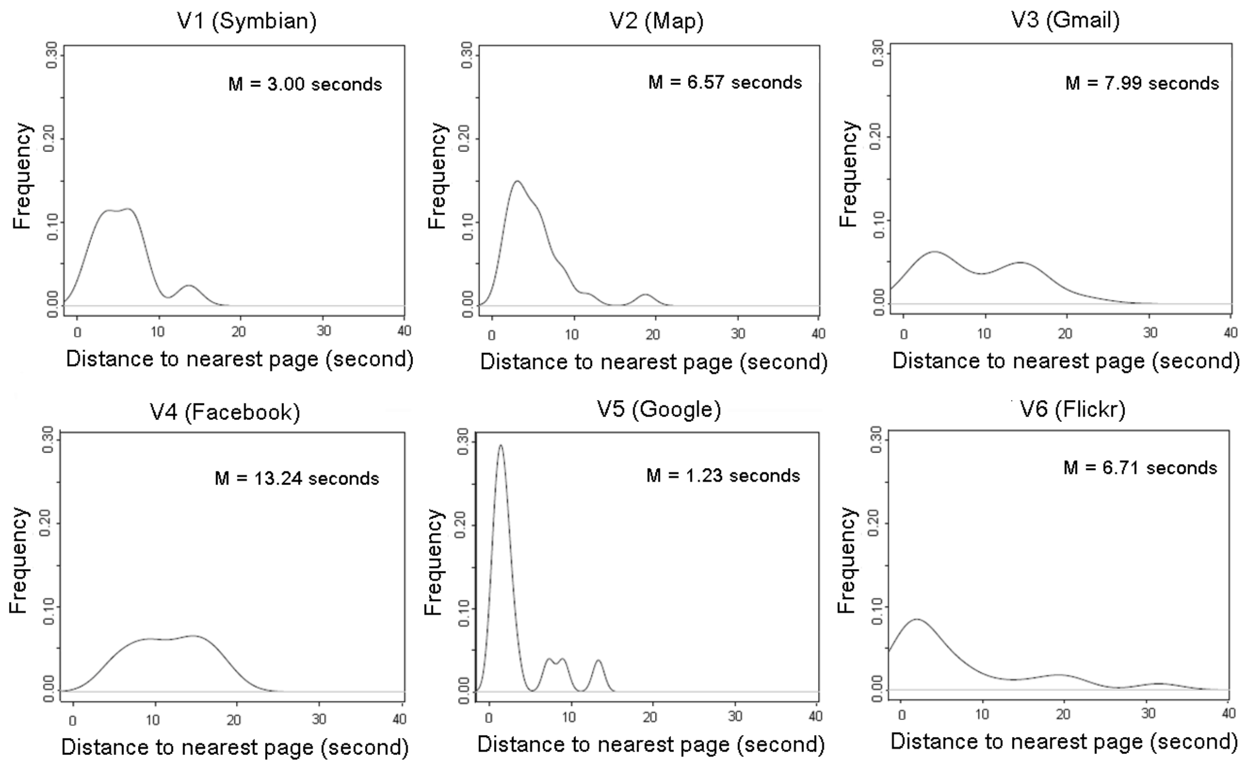

V1 (Symbian)
M = 3.00 seconds
V2 (Map)
M = 6.57 seconds
V3 (Gmail)
M = 7.99 seconds
V4 (Facebook)
M = 13.24 seconds
V5 (Google)
M = 1.23 seconds
V6 (Flickr)
M = 6.71 seconds
Frequency
0.00
0.10
0.20
0.30
0
10
20
30
40
Distance to nearest page (second)

Figure 6.

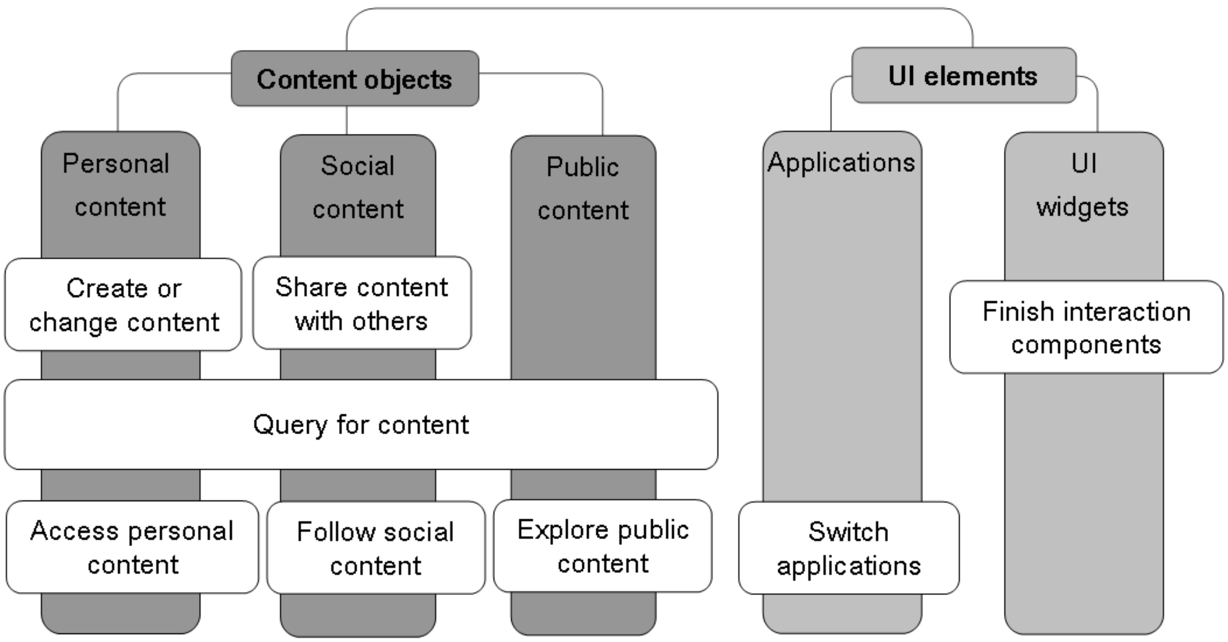

Content objects
UI elements
Personal content
Social content
Public content
Applications
UI widgets
Create or change content
Share content with others
Finish interaction components
Query for content
Access personal content
Follow social content
Explore public content
Switch applications

Figure 7.

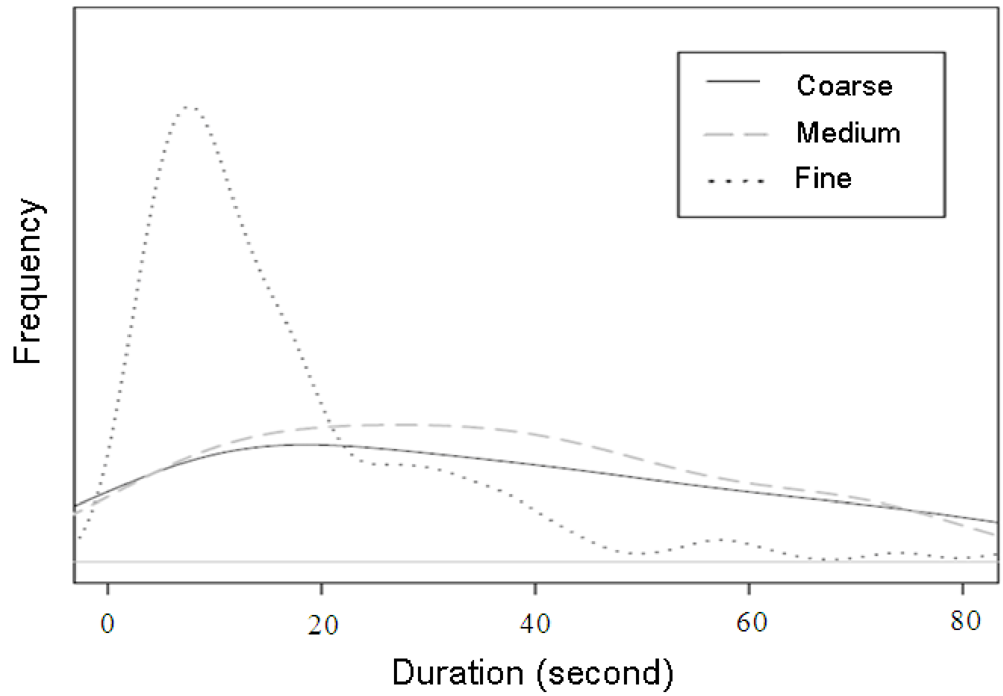
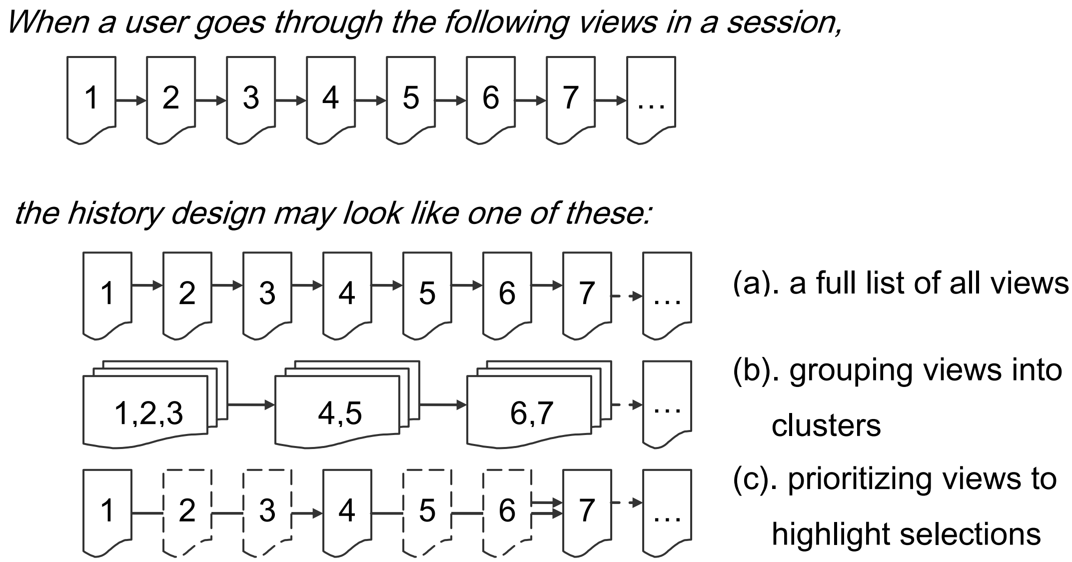


Figure 8.


When a user goes through the following views in a session,
1
2
3
4
5
6
7
…
the history design may look like one of these:
1
2
3
4
5
6
7
…
(a). a full list of all views
1,2,3
4,5
6,7
…
(b). grouping views into clusters
1
2
3
4
5
6
7
…
(c). prioritizing views to highlight selections