Identifying Phenotypes & Genotypes: A Case Study Evaluating an In-car Navigation System

Georgios Papatzanis[1], Paul Curzon[1], and Ann Blandford[2]

[1] Department of Computer Science, Queen Mary, University of London,
Mile End Road, London, E1 4NS, UK
gp@dcs.qmul.ac.uk, pc@dcs.qmul.ac.uk
[2] UCL Interaction Centre, University College London,
Remax House, 31-32 Alfred Place, London WC1E 7DP, UK
A.Blandford@ucl.ac.uk

Abstract. There are a range of different usability evaluation methods: both analytical and empirical. The appropriate choice is not always clear, especially for new technologies. In-car navigation systems are an example of how multimodal technologies are increasingly becoming part of our everyday life. Their usability is important, as badly designed systems can induce errors resulting in situations where driving safety may be compromised. In this paper we use a study on the usability of a navigation device when the user is setting set up an itinerary to investigate the scope of different classes of approach. Four analytical and one empirical techniques were used to evaluate the usability of the device. We analyse the results produced by the two classes of approach – analytical versus empirical – and compare them in terms of their diversity and the insight they provide to the analyst in respect to the overall usability of the system and its potential improvement. Results suggest a link between genotypes and the analytical class of approach and phenotypes in the empirical class of approach. We also illustrate how the classes of approach complement each other, providing a greater insight into the usability of a system.

Keywords: Usability evaluation; UEMs; In-car navigation; Cognitive Walkthrough; UAN; EMU; Design Criteria; Phenotypes; Genotypes.

1 Introduction

Various techniques can be used for the evaluation of interactive systems. Techniques are classified according to their approaches in conducting evaluation: *analytically* when a simulation is performed by an expert/analyst to predict the behaviour of the user and detect potential problems, without the involvement of users; *empirically* when the system is tested by users while their performance and problems are recorded.

The aim of this paper is to report on a study where both analytical and empirical approaches were employed to evaluate an in-car navigation device. In this study we concentrated solely on tasks related to the programming of the device (destination entry) before the user starts driving the car. We look at the results from a qualitative perspective; we do not seek to establish efficiency counts (number of usability problems) for different techniques or approaches. Instead we analyse the results produced by the two classes of approach – analytical and empirical – and compare them in terms of their diversity and the insight they provide to the analyst in respect of the overall usability of the system and its potential improvement. We investigate the variance of results between the classes of approach and explore the association of genotypes and phenotypes with the empirical and analytical classes of approach respectively.

2 Background

Car navigation systems are designed to guide the driver through a generated route toward a selected destination. Drivers are interacting with such devices when programming the destination point and customising the route (touch screens, dials, voice input) and whilst driving when receiving instructions from the device (maps, visual cues, voice instructions). As a result navigation systems can cause driver distraction (a) when entering information in the device and (b) when following the driving instructions issued by the system. The different modes of interaction with the device have varying effects on the driving performance.

The usability of navigation devices is a contributing factor to the overall safety of car driving. Nowakowski et al [19] carried out heuristic analysis and user testing on navigation systems and exposed a set of recurring usability problems. Nowakowski *et al* identified problems in both destination entry and guidance modes: (a) layout and labelling of the control menus; audio and visual feedback; order of entry of destination information, and (b) starting guidance and ending; display design and readability; voice guidance and timing; rerouting. In this paper we examine aspects of the device related to the preparation of a route, before the device commences with the navigational instructions to the car driver.

Various case studies are reported in the literature with regard to the evaluation of usability methods ([18], [1], [12], [4], [6]). Comparisons between methods have been carried out in terms of problem count, scope, validity, evaluator effect, etc. Wright and Monk [22] also carried out case studies reporting on the difference of usability evaluation results obtained between users or usability experts and the system designers when applying cooperative evaluation to the same system.

In this study we take a different perspective and make a comparison between analytical and empirical classes of approach on two discrete dimensions. The first dimension considers usability problems identified and the insight they provide to the analyst into the usability of a system. Secondly, we look at the usability issues in terms of phenotypes – overt and observable manifestations of an incorrectly performed action – and the contrasting genotype – the underlying likely cause which eventually can account for the phenotype [8] [9]. Phenotypes describe observable behaviour, while genotypes are concerned with the interpretation of such behaviour.

3 Method

The case study was executed in two discrete parts: analytical and empirical evaluation, followed by an analysis of the results comparing the two classes of approach. In each, the usability of the selected application was assessed against a predefined scenario and set of tasks (see Table 1). The scenario and tasks were based on the activities carried out by the driver prior to driving to the destination, i.e., preparing an itinerary on the navigation device. Such tasks take place in the car while stationary. This set of tasks enabled us to assess a wide range of primary functions that are frequently used in such devices.

Table 1. Sample tasks used for the evaluation

Task 1: Program the device to reach the city centre of Leeds.

Task 2: Program the device to reach the following restaurant before the final destination.

World Service
Newdigate House
Castle Gate
Nottingham
NG1 6AF

Task 3: Check the route to see if you are using the M621. If you do, program the device to avoid this part of the route.

In the analytical part of the study we applied a series of analytical methods in the evaluation of the navigation system. The first author of the paper carried out the analytical evaluations of the system. The personal judgement and experience of an analyst, may have a significant impact on the results (known as the evaluator effect [6] or craft skill). Nevertheless, in this study we focus more on the types of problems reported by each class of approach, rather than contrasting problem counts. We compare the results as identified by the different classes of approach, empirical vs. analytical, rather than comparing the different sets of issues within each class of approach. As a result, the evaluator effect has minimal impact on our comparison. Furthermore, two usability experts independently reanalysed the results with respect to genotypes and phenotypes.

Four methods were chosen for the analytical part of the study, employing a diverse approach to evaluation. In the subsequent sections, we describe these techniques and identify various issues pertaining both to their applicability and their effectiveness as identified during the study. The methods selected are characterised by a varying degree of formality, with each advocating a different approach to user interface evaluation. Each method has its own potential merits in the evaluation of this device. Cognitive Walkthrough [20] was selected as it is most suitable for walk-up-and-use interfaces. EMU (Evaluating Multi-Modal Usability) [10] is a technique specifically implemented for multimodal systems, thus appropriate for this type of device. UAN (User Action Notation) [5] provides an extensive notation, incorporating temporal issues and patterns for the specification of the interface. Leveson's design guidelines [13] were selected because of their focus on error detection and analysis. The diversity of these techniques gives us an increased capacity for the detection of usability issues, giving a wide range to compare against those found empirically.

In the second part of the study we carried out an empirical evaluation of the device, using the same scenario and tasks as in the first part of the study. The empirical evaluation was carried out in a usability laboratory, as the context (being in a car) is not relevant for the set of tasks selected for this study. We focused our attention on the usability issues that drivers encounter in the use of such devices, employing an exploratory approach.

3.1 Car Navigation System

Navigation systems are increasingly becoming standard equipment in motor vehicles. Their usability is an important factor, as badly designed systems can induce errors resulting in situations where driving safety is compromised. Although manufacturers suggest that users must read the entire manual before operating such navigation systems, it is often the case that they are used by drivers as walk-up-and-use devices.

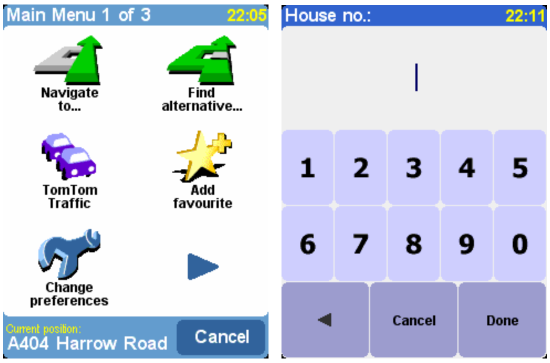

Fig. 1. (a) Main menu & (b) House number entry

The navigational device selected for this study utilises the TomTom Navigator 5 application running on an HP iPAQ handheld computer. The user can manipulate the application through the user interface displayed on the touch screen of the device. The device offers visual and voice instructions to the user in order to guide them through the itinerary. The system offers the functionality usually found in navigational devices, such as looking up and navigating to an address or point of interest, re-routing and generating different routes to a selected destination. The system is accessed via a touch-screen based interface comprising a set of menus and data entry screens (see Fig. 1).

4 Analytical Techniques

In this section, we briefly outline each analytical method used, describing the empirical study methodology in the next section.

4.1 Cognitive Walkthrough

Cognitive Walkthrough (CW) is an informal inspection methodology for systematically evaluating features of an interface in the context of the exploratory theory CE+ [14] [20]. Wharton *et al* [20] present CW as a theoretically structured evaluation process that follows the application of a set of questions asked about each step in the task, derived from the underlying theory, and attempting to focus the attention of the analyst on the CE+ claims. The questions are preceded by a task analysis and the selection of the appropriate sequence of user actions to successfully perform a task (preparatory phase). During the execution of the method (analysis phase), the analyst simulates the execution of the sequence of user actions and assesses the ease of learning of the design, by using the questions as summarised and exemplified in Table 2.

Table 2. Cognitive Walkthrough extract

*Task***:** Enter house number

Question 1: Will the users try to achieve the right effect?

No. The system requires information not known by the driver.

Question 2: Will the user notice the correct action is available?

Probably not. The driver should select 'done' on this screen, in order to avoid inputting a house number.

Question 3: Will the user associate the correct action with the effect trying to be achieved?

No. The driver might attempt to enter a random number to skip this screen.

Question 4: If the correct action is performed, will the user see that progress is being made towards the solution of the task?

Not really. Once the selection is made the system automatically starts calculating the route without any further confirmation. The markers and labels on the map are indiscernible or non-existent and cannot confirm the route that the driver has been trying to build up.

In this extract the user is asked to enter a house number as part of the destination input, although such information is not provided in the use scenario. Although this information is not required by the system, there is no clear way to skip this step.

In this study CW reported a series of issues relating to feedback, consistency of design, labels, task structure, and user interface navigation. The bulk of the issues identified by the technique are attributed to the analyst’s craft skill, rather than the technique itself. Nevertheless, the technique led the analyst to engage deeply with the system in order to arrive at these results.

4.2 UAN (User Action Notation)

UAN [5] [7] is a behaviour-based notation specifying user actions, computer feedback and interface internal state at the same time. UAN is primarily a shorthand way to represent steps (such as “mouse down”) that a user would take to perform a task on a given user interface, existing or under development.

The notation is semi-formal in that it makes use of visually *onomatopoeic* symbols, for example *Mv* represents a “mouse down” action, while *M^* represents a “mouse up” action. The goal of UAN is to represent simple and complex user tasks in a notation, that is easy to read and write, but one that is more formal, clear, precise, and unambiguous than English prose. As it is not overly formal it is assumed that designers can learn the notation without major problems.

Table 3. Extract from UAN specification of the user interface

TASK: **Enter street number**			
USER ACTIONS	**INTERFACE FEEDBACK**	**INTERFACE STATE**	**CONNECTION TO COMPUTATION**
&~ [number’]* Mv	number’!	key selected = number’	put number’in field
M^	number’-!	key selected = null	
~ [Done] Mv	Done!		
M^	Done-!		If field isNull then number = default else selected number = field number; go to map screen

In the extract shown in Table 3, we describe the interaction with the user interface in the house entry dialogue. During the interaction the user selects the appropriate number (~ [number’]*) using the virtual numerical keyboard, while the ‘done’ button is used to complete the task. From the specification we can easily distinguish the feedback (number’!) provided at each step of the interaction, as the system updates (key selected = number’) its variables and displays (put number’in field) the relevant information.

Although UAN is not necessarily a suitable technique for identifying usability problems, the specification of the interface enforces the analyst to deconstruct the user interface and identify issues hidden in its design. In the UAN analysis we identified mainly issues related to feedback, design and labelling of the interface.

4.3 EMU - Evaluation Multi-Modal Usability

EMU (Evaluation Multi-modal Usability) [10] is a methodology developed to address specifically the evaluation of usability of multi-modal systems. The scope of the methodology extends to issues related to user knowledge and use in context, with a special focus on the issues concerned with the physical relationship between the user and the device [1]. Multimodal interfaces can enhance the user's understanding, as they provide more communication channels between the user and the system.

Table 4. EMU stages

Stage 1. Define the task that is to be analysed
Stage 2. Modality lists
Stage 3. Define the user, system and environment variables
Stage 4. Profiles compared to modality listings
Stage 5. Interaction modality listing
Stage 6. Add in clashes, etc.
Stage 7. Assess the use of modalities
Stage 8. Final report.

EMU methodology presents a novel approach to the evaluation of multimodal systems. It presents a comprehensive taxonomy, underpinned by a new theory on multimodality, tightly coupled with a notational representation and a structured step-by-step approach for its application. In this evaluation, we applied the methodology as described in the EMU tutorial [10]. The methodology is executed in several stages (see Table 4) in order to identify the various modalities (see Table 5) of the interaction and any usability issues resulting from these modalities.

Table 5. Extract from EMU analysis

Display
[UE hap-sym-dis]
*user types the house number *
[SR hap-sym-dis]
*system records house number *
[SE vis-sym-dis]
system flashes pressed buttons
[UR vis-sym-dis]
user sees pressed button
and
[SE vis-lex-cont]
number appears on house number field
[UR vis-lex-cont]
user reads house number field
precon: UE [hap-sym-dis]
user types numbers

key
SE: System Expressive (expressed by the system)
SR: System Receptive (received by the system)
UE: User Expressive (expressed by the user)
UR: User Receptive (received by the user)
hap: haptic, vis: visual, lex: lexical,
sym: symbolic, dis: discrete, cont: continuous

Table 5 gives us an extract of the EMU analysis for the house entry dialogue of the system as shown in Fig.1 (b). The user enters the information ([UE hap-sym-dis]) into the system using the touch screen display. As the system receives the information ([SR hap-sym-dis]), the appropriate visual feedback ([SE vis-sym-dis]) is received by the user ([UR vis-sym-dis]) for each button pressed. At the same time the user can read the information ([UR vis-lex-cont]) provided to the system, as it is shown in the relevant display ([SE vis-lex-cont]).

Due to the nature of the tasks under evaluation, there were only a very limited number of modality clashes identified as part of EMU analysis. Nevertheless, the analysis gave the analyst the opportunity to examine the system from a different perspective, resulting in an extensive set of usability problems, with a wider scope not solely related to multimodal issues, but also labelling, interface design, and interface navigation issues.

4.4 Design Guidelines

The use of design guidelines (DG) or design criteria has been a common practice for the evaluation of user interfaces. The conformance of the interface design to an appropriate set of guidelines can improve the usability of an application. In the HCI literature one can find different sets of guidelines to suit different domains and applications [15]. Guidelines can be used for helping designers resolve design problems, or for the evaluation of an interface.

The design guidelines of Nielsen and Molich [17] have been widely used in the HCI community in order to improve the usability of interactive systems. This method, heuristic evaluation [17], is suitable for quick and relatively easy evaluation. The analyst carries out a systematic inspection of the interface to identify usability problems against a set of guidelines, also known as heuristics.

In the field of safety-critical systems, the analyst seeks to identify high-risk tasks and potentially safety-critical user errors through system hazard analysis. Various sets of guidelines for detecting design flaws, which might cause errors leading to safety-critical situations, can be found in the literature (e.g., [11] [13]) .

Jaffe [11] and Leveson [13] have created sets of guidelines for the design of safety-critical systems. In this study, we used a subset of the Human-Machine Interface (HMI) Guidelines [13]. These guidelines are based partly on an underlying mathematical model, but to a greater extent on the experience of the authors in the design and evaluation of safety-critical systems used in cockpits. As a result these guidelines are greatly influenced by issues pertinent to the particular domain.

Although these guidelines were not intended for the usability evaluation of interactive systems, we applied them in a similar way that an analyst would apply guidelines in Heuristic Evaluation [16]. Every screen of the system in the task sequence was assessed against a subset of the Design Guidelines. During the evaluation we only used a restricted subset as many of them were either domain-specific or irrelevant to our device. This reduced the number of times that the analyst had to traverse through the list of guidelines during the evaluation session.

Table 6. Extract from the subset of Design Guidelines used for the evaluation of the interface

Design Guidelines

1. Design for error tolerance: (a) make errors observable, (b) provide time to reverse them, and (c) provide compensating actions
2. Design to stereotypes and cultural norms
3. Provide adequate feedback to keep operators in the loop.
4. Provide facilities for operators to experiment, to update their mental models, and to learn about the system. Design to enhance the operator's ability to make decisions and to intervene when required in emergencies.
5. Do not overload the operator with too much information. Provide ways for the operator to get additional information that the designer did not foresee would be needed in a particular situation.
6. Design to aid the operator, not take over.

Applying this technique in the house entry dialogue (Fig. 1 (a)), as shown before with other techniques, we identified several issues that violated the design guidelines. Table 7 gives extracts from the analysis detailing some of the problems and the associated guidelines that have been violated.

Table 7. Extract from the Design Guidelines analysis of the system

(Guideline 1) If the users change their mind or realise they needed a different postcode, it is impossible to return to the previous page to rectify their action. The user will have to cancel the interaction and start again from Step 1.

(Guideline 2) It is not possible on this page to confirm that the right selection has been made in the previous page. An instant flashing message is displayed to the user when the page loads, but it can be easily missed.

(Guideline 3) There is no label associated with the arrow button.

DG were drafted to be used for the design of safety-critical systems. In this study, we identified a range of usability problems in the process of analysis – labelling, navigation, feedback, as well as issues relating to error recovery which are specifically targeted by the method.

5 Empirical Study

An empirical study can potentially give important insights regarding issues of context of use that analytical methods might fail to capture. We also investigated issues concerning whether analytic and empirical methods can be combined into a composite method for comprehensive coverage of problems in such environments. It has previously been suggested [2] that the combination of empirical and theoretical analysis can provide a greater insight than the individual approaches into the issues of such an application area.

Eight users participated in the experiment, including both male and female members of the academic community. The empirical study was split into two parts. All participants participated in both parts of the experiment. The first part was a training session where users were expected to follow a given scenario and carry out a set of three tasks (see table 8). During this part of the trial, the users were allowed to ask questions of the experimenter. The goal of this session was to allow the participants to familiarise themselves with the device before continuing to the main trial session. Participants were provided with a sheet containing the scenario and tasks and a print-out containing a set of screens from the device.

Table 8. Tasks used for training session

Task 1: Program the device to reach the Berkeley Hotel, Brighton.

Task 2: Program the device to reach the following restaurant before the final destination.

IKEA Croydon
Volta Way
Croydon
CR0 4UZ

Task 3: Check your route to see if you are using A22. If you are, program the device to avoid this part of the route.

During the second part, users followed a different set of (similar) tasks in a new scenario. At this stage the experimenter did not interfere with the tasks. In the second part of the empirical study we used the task list that was also used with the analytical techniques of the study (see Table 1). In both sessions of the experiment we used TomTom Navigator 5 software running on an iPAQ handheld computer connected to a TomTom GPS device via Bluetooth, as described in previous sections.

During the experimental trials we collected video and audio data from the interaction between the user and the system using a video camera. We also captured a video stream of the information shown on the screen of the iPAQ device. The video data for each participant were synchronised and merged before we started a thorough analysis of the interaction.

Firstly, we started with the transcription of the sequence of actions that each user followed in order to achieve the tasks as set out in the experiment trials. Each interaction step was recorded and matched against the current state of the interaction device. Having completed this process, we analysed the data, in order to understand the problems that the users encountered during the interaction and how their sequence of actions compare to the sequence of actions required to successfully complete the tasks. We grouped together relevant sequence events and identified repeating patterns between users in their interactions.

6 Results from Analytical and Empirical Evaluations

In this study we examined several parts of the system over three different tasks selected for evaluation. We identified a set of over 50 usability problems attributed to one or more techniques.

Although the analytical and empirical techniques managed to identify a similar number of issues during the analysis, the types of issues varied significantly. Each *class of approach* identified a distinct set of issues, while only a few usability problems were identified by both classes of approach.

We overview briefly below the two subsets of usability problems – analytical and empirical – and where these subsets intersect. For the purposes of the analysis we give for illustration (Table 9) a representative sample of usability problems collated during the analytical and empirical study.

Table 9. Extract of usability problems list

		CW	UAN	EMU	DG	EMP
1.	No way to edit the route from view menu			☒	☒	☒
2.	No clear way to bypass house number entry	☒		☒	☒	☒
3.	Invalid input through address entry					☒
4.	Wrong mode					☒
5.	Inappropriate headings in menus	☒	☒	☒	☒	
6.	Inconsistent colour scheme within menus	☒			☒	

The first two usability problems identified in the Table 9 were captured by both analytical and empirical techniques.

1. No way to edit the route from view menu

The design of the menu structure prohibited the users from making any changes to the generated route from the set of menus used to view the route. As a result, users frustratingly navigated through the system menus to locate the appropriate functionality.

2. No clear way to bypass house number entry

The second task of the user trial involved programming the navigation device to reach a restaurant. The address provided to the user in the context of this trial did not include a street number. Nevertheless, the system asks for this piece of information (see Fig.1 (b)) as part of the address, without an obvious way to bypass it.

Analytical techniques identified the issue in the study, offering design recommendations to resolve it. In the empirical study, as users were not aware of a street number for the restaurant, they employed various strategies, as there was no evident way to skip this step. Some users entered a random number, while some others chose the 'crossing menu' functionality in order to complete their task.

The next two usability problems (number 3 & 4) identified were captured by the empirical evaluation only.

3. Invalid input through address entry

The system offers different interactive dialogues to the user to input the information about the destination in terms of an address, postcode, point of interest, city centre, etc. Users repeatedly attempted to input through the address dialogue, information other than that asked for at the time by the device. Although the first screen of this dialogue asks for the city name of the destination, users tried to enter the postcode, name of the restaurant, street name, etc. Apparently users did not identify the specific dialogue for postcode entry, subsequently trying to communicate the postcode to the system through this menu, since another option was not readily available. This led to confusion, decreasing significantly the usability of the system.

4. Wrong mode

Another issue identified in the empirical study only refers to the user being in the wrong mode. The system incorrectly captured the intentions of the user without the user realising. This was identified as the user attempted to carry out the 2nd task, i.e., setting up an intermediate destination in the itinerary, through the address dialogue. More specifically, the user input part of the name (N-O-T-T) of the stopover town (Nottingham) and the system automatically updated the appropriate list, according to the user input. As it was being updated, Nottingham momentarily appeared on top of the list, before it went to second place in order to be replaced by Notting Hill. Nevertheless, the user selected the first item on the list, having not realised that the list had changed before the selection was made.

Under these circumstances the user arrived at the next screen, 'Travel via/Address/Street entry', under the illusion that Nottingham was selected in the previous screen. As a result the user was unsuccessful in locating the street or the restaurant on the list, as the wrong city was selected.

The last two usability problems identified that we discuss here were captured only by analytical class of approach:

5. Inappropriate headings in menus

The lack of appropriate headings throughout the application was picked up by all analytical techniques applied in this study. Titles are necessary as they provide orientation and constant feedback to the user. Missing or inappropriately used headings decrease significantly the usability of a system.

6. Inconsistent colour scheme within menus

Colour schemes can be used to group together similar functions, while at the same time offering the sense of continuity to the user when progressing through the task. In this system the colour scheme is used inconsistently, resulting in inappropriate feedback and sense of confusion by the user. This was picked up by DG as it violated the respective guideline, while it was also identified in the process of the CW.

7 Analysis of the results

The types of issues captured by analytical and empirical techniques vary significantly. Some usability problems were identified by both classes of approach (analytical and empirical), but many were identified only by one or the other. In the previous section we presented a set of usability problems representing these categories and as tokens of the usability problems identified.

One important aspect that emerges when looking at the results is the variability between the coverage of results reported by analytical and empirical approaches. There is only a small overlap on the issues identified by the two approaches. The vast majority of usability problems were independently identified by one class of approach only.

Under closer investigation we also observe that the type of problems detected by the approaches is significant. While the analytical techniques identified mainly usability problems that might create difficulties to the users, the empirical data demonstrated specific instances of user behaviour where users experienced such difficulties. The usability problems reported by the empirical approach are associated with the manifestations of user errors, while the usability problems reported by the analytical approach correspond to the underlying cause of such manifestations. This correspondence thus relates to the phenotype – observable manifestations of an incorrectly performed action – and the contrasting genotype – the underlying likely cause [8] [9].

Table 10. Extract of reanalysis of usability problems

		Analytical	Empirical	
1.	No way to edit the route from view menu	☒	☒	*genotype*
2.	No clear way to bypass house number entry	☒	☒	*genotype*
3.	Invalid input through address entry		☒	*phenotype*
4.	Wrong mode		☒	*phenotype*
5.	Inappropriate headings in menus	☒		*genotype*
6.	Inconsistent colour scheme within menus	☒		*genotype*

In order to investigate the association of genotypes and phenotypes with their respective classes of approach – empirical and analytical, the first author and a further two usability experts independently assessed the issues identified in the study in terms of genotypes and phenotypes. The experts did not have any prior knowledge of the results or their association to any technique or class of approach. They were provided with the complete list of issues, as identified by both classes, and were instructed to assign each issue as a phenotype or as a genotype. The experts were able to match the majority (over 95%) of the issues to the type of error, as we had hypothesised with the correlation between genotypes, phenotypes and their respective classes of approach.

In Table 10 you can examine the reanalysis of the usability problems presented in section 6. More specifically, the issues identified by the empirical class of approach were assigned as genotypes, whereas the issues identified by the analytical class of approach were assigned as phenotypes. In the extract presented in Table 10, the problems captured by both classes of approach were classified as genotypes.

Matching phenotypes to their respective genotypes during the analysis of the results in the study turned out to be a difficult feat. Although we were able to identify several manifestations of user difficulties, we were unable to directly pinpoint the underlying cause; we could only theorise about possible candidates. For example, a phenotype identified in the study was issue three from Table 9. As explained in section 6, the user attempted to make an invalid entry through the address dialogue. There are several genotypes that can be potentially associated with this issue, such as inappropriate headings, inconsistent interface design, grouping of functions, etc. Although some of them could be perspective candidates it is not easy to establish a link between them. The lack of the specific underlying causes prevents us from making design recommendations in order to remove such user difficulties, identified as phenotypes. Such relationships, between genotypes and phenotypes could eventually be established through further experimental studies examining the appearance (or not) of the phenotypes, once a genotype has been removed from the system. However this would be a very time-consuming approach.

Usability evaluation methods are used in order to improve the usability of a system. This is done through the identification of usability problems and a set of design recommendations, which are subsequently applied to improve the usability of the system under investigation. We have seen in this study that the empirical study mainly focused on the identification of phenotypes, which does not lead directly to the improvement of a system, as it does not provide causal explanations needed in many cases as a precursor for recommendations on how to do so. Nevertheless, the phenotypes also serve their purpose as they are reminders to designers and developers of the difficulties or problems encountered by the users and their satisfaction while using the system.

Although an empirical approach can identify in a system difficulty of use or usability problems, it does not readily identify or resolve the underlying causes of the issues identified. An alternative approach should be followed for the identification of the genotypes. As demonstrated in this study, the analytical approaches fare well in this task. The coverage of results collated by the analytical techniques used in this study concentrates mainly on the genotypes. Furthermore an explicit part of some of the techniques – such as EMU and CW – is the identification of design recommendations that can be used for eradicating the genotypes from the system under evaluation.

Nevertheless, this does not reduce the value of the empirical approach. Wixon [21] argues that the evaluation of a system is best accomplished within the context of use for each system, advocating a more exploratory approach, through the use of case studies and user involvement. Furniss [3] also argues that demonstrating user problems (phenotypes) is a significant step for persuading design teams and relevant stakeholders to introduce changes to systems. In contrast, expert reports (describing genotypes) can be more easily dismissed. Thus, the use of phenotypes might be used for persuading the appropriate stakeholders as needed, while genotypes can help the design times understand better the underlying causes and offer more appropriate solutions.

As illustrated above neither of the two approaches can serve as a panacea for evaluating an interactive system. Using an analytical or empirical approach can only have a limited effect on the overall usability of the system. Each approach offers different insights and power to the analyst and the synergy of a combined approach can provide a more complete approach to usability evaluation.

8 Conclusion

In this study we set out to compare different evaluation techniques by evaluating the usability of a car navigation device. Our efforts were concentrated on the aspects of the device relating to the preparation of a route, before the device commences with the navigational instructions to the driver of the car.

In the previous sections we examined the analytical and empirical techniques that were employed during the study. Each technique employed in this study offers a different perspective into the usability evaluation of interactive systems and identified different sets of issues. In this study we focused on the kind of usability problems reported from each class of approach. According to the results of the study, the analytical class of approach is most powerful as a way of identifying genotypes, while the empirical class of approach is best at identifying phenotypes. These results support the argument that a combination of analytical and empirical approaches can offer a richer insight into the usability of the system and give the usability practitioner greater argumentative power, as their findings complement each other.

The combinatory use of the complementary approaches described above still remains a challenge for the analyst. The association of phenotypes with their respective genotypes is a difficult task, but necessary in the process of increasing the usability of a system, when adopting such an approach. Further work needs to be carried out into making this process easier for the analyst to undertake. Taxonomies identifying domain specific genotypes and phenotypes could eventually assist the analyst relating observational behaviour to underlying cause, resulting in a deeper insight into the usability of a system.

In order to assess further the scope of each technique and approach in a dynamic environment, we are carrying out another study where the tasks selected are representative of the user experience while driving and taking instructions from a navigation device. This future study will give us further insight into the appropriateness of the methods when using such devices in a constantly changing environment and where the goals of the users are not preconceived as is the case in this study.

Acknowledgments. The work described in this paper has been funded by the EPSRC Human Error Modelling project (GR/S67494/01 and GR/S67500/01). We are grateful to all participants in our studies.

References

1. Blandford, A.E., Hyde, J.K., Connell, I., and Green, T.R.G., *Scoping Analytical Usability Evaluation Methods: a Case Study*. Submitted for journal publication, 2006.
2. Curzon, P., Blandford, A., Butterworth, R., and Bhogal, R. *Interaction Design Issues for Car Navigation Systems*. in *16th British HCI Conference*. 2002: BCS.
3. Furniss, D., Blandford, A., and Curzon, P., *Usability Work in Professional Website Design: Insights from Practitioners' Perspectives*, in *Maturing Usability: Quality in Software, Interaction and Value*, Law, E., Hvannberg, E., and Cockton, G., Editors. (forthcoming), Springer.
4. Gray, W.D. and Salzman, M.C., *Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods.* 1998. **13**(3): p. 203-261.
5. Hartson, H.R. and Gray, P.D., *Temporal Aspects of Tasks in the User Action Notation.* Human-Computer Interaction, 1992. **7**(1): p. 1-45.
6. Hertzum, M. and Jacobsen, N.E., *The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods.* 2003. **15**(1): p. 183-204.
7. Hix, D. and Hartson, H.R., *Developing User Interfaces: Ensuring Usability Through Product and Process*. 1993: John Wiley and Sons.
8. Hollnagel, E., *The phenotype of erroneous actions.* International Journal of Man-Machine Studies, 1993. **39**(1): p. 1-32.
9. Hollnagel, E., *Cognitive Reliability and Error Analysis Method*. 1998, Oxford: Elsevier Science Ltd.
10. Hyde, J.K., *Multi-Modal Usability Evaluation*. 2002, PhD thesis. Middlesex University.
11. Jaffe, M.S., Leveson, N.G., Heimdahl, M.P.E., and Melhart, B.E., *Software Requirements Analysis for Real-Time Process-Control Systems.* IEEE Trans. on Software Engineering 1991. **17**(3): p. 241-258.
12. Jeffries, R., Miller, J.R., Wharton, C., and Uyeda, K., *User interface evaluation in the real world: a comparison of four techniques*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*. 1991, ACM Press: New Orleans, Louisiana, United States.
13. Leveson, N.G., *Safeware: System Safety and Computers*. 1995: Addison-Wesley.
14. Lewis, C., Polson, P.G., Wharton, C., and Rieman, J., *Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*. 1990, ACM Press: Seattle, Washington, United States.
15. Newman, W.M., Lamming, M.G., and Lamming, M., *Interactive System Design*. 1995: Addison-Wesley Longman Publishing Co., Inc. 468.
16. Nielsen, J., *Heuristic Evaluation*, in *Usability Inspection Methods* Nielsen, J. and Mack, R.L., Editors. 1994, John Wiley & Sons: New York, NY.

17. Nielsen, J. and Molich, R., *Heuristic evaluation of user interfaces*, in *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*. 1990, ACM Press: Seattle, Washington, United States.
18. Nørgaard, M. and Hornbæk, K., *What do usability evaluators do in practice?: an explorative study of think-aloud testing*, in *Proceedings of the 6th ACM conference on Designing Interactive systems*. 2006, ACM Press: University Park, PA, USA.
19. Nowakowski, C., Green, P., and Tsimhoni, O. , *Common Automotive Navigation System Usability Problems and a Standard Test Protocol to Identify Them*, in *ITS-America 2003 Annual Meeting*. 2003, Intelligent Transportation Society of America: Washington, D.C.
20. Wharton, C., Rieman, J., Lewis, C., and Polson, P., *The cognitive walkthrough method: a practitioner's guide*, in *Usability inspection methods*. 1994, John Wiley & Sons, Inc. p. 105-140.
21. Wixon, D., *Evaluating usability methods: why the current literature fails the practitioner.* interactions, 2003. **10**(4): p. 28-34.
22. Wright, P.C. and Monk, A.F., *A cost-effective evaluation method for use by designers.* 1991. **35**(6): p. 891-912.