A Comprehensive Learnability Evaluation Method for In-Car Navigation Devices

Elliott Noel, Blair Nonnecke, Lana Trick
University of Guelph, Funded by Auto21

ABSTRACT

Although usability is a broad subject, the literature relating to in-car navigation devices tends to focus primarily on efficiency of use. In this paper, we explore the neglected issue of learnability, that is the ease with which users learn to use, as well as some justification of its importance as a distinct issue.

We present an evaluation method designed to identify learnability related problems with in-car navigation devices, along with a complete set of basic tasks with which to apply it to. Our method is applied to a device, identifying various learnability problems. Recommendations for builders of in-car systems are made based on these observations.

INTRODUCTION

As fervent global competition and the law of diminishing returns squeezes automaker profitability, new value-added products and services are growing in significance. In-car navigation devices are one such product being positioned for this purpose. With only one in five interested consumers currently possessing one [1], the untapped potential is apparent. If this market is to be properly developed, these devices must be usable, ergonomic. It is critical to their adoption that consumer expectations be properly met.

While “usability” is subject to a number of definitions, the basic principles remain invariant. According to Nielsen [10], it is characterized by efficiency of use (sometimes called “throughput”), rate of use error, user satisfaction, learnability, and memorizability. For in-car devices, efficiency is intimately tied to safety – attention expended on the device is attention taken from driving. For this reason, some research has been conducted focusing primarily on this particular facet. Green and Burnett [9], for example, define usability as effectiveness, efficiency, and satisfaction – no mention is made of learnability. In these types of studies, the time taken for a task is often used as a performance metric, along with error rates.

Learnability and memorizability, on the other hand, have been largely neglected. We contend that this is a significant oversight, as learnability is a critical factor for the adoption of new technology. A user, unable to make use of an expensive device with high expectations within a reasonable time period, is likely to become dissatisfied. When one considers important markets such as rental vehicles, where the driver is likely to be using the particular system for the first time and likely to be in an unfamiliar locale, learnability becomes significant for the device’s success.

Memorizability is similarly overlooked, but equally important. A study on navigation devices by Metrinomics [1] reports that “usage is limited to those situations where the driver is uncertain about their location or their destination”. The study mentions that 64% of users use the device as a back-up for unfamiliar routes. Thus despite infrequency of use, it is imperative that the device be conducive to the user remembering how to operate it when called upon. In a sense, the device must facilitate “re-learning” its interface with ease.

Learnability is also intimately tied to safety, with obvious regulatory consequences. Research focusing on efficiency without taking into account frequency of use and level of familiarity inadequately addresses this issue.

The purpose of our research, then, is to provide those building these systems with a comprehensive method for evaluating their learnability and memorizability. Our goal is to improve user satisfaction and safety, as well as reduce barriers to their widespread adoption.

METHODOLOGY

TASKS

A fundamental principle of our evaluation of navigation devices is the notion that real users are to be observed performing realistic tasks, rather than making use solely of expert analysis. In order to accomplish this, a meaningful and complete set of user tasks must be
identified for our subjects to perform. We identify these here for use in our study.

The two most significant uses of a navigation system are, according to market research [1], orientation and route planning. Orientation is the simpler of the two – a user is unsure as to his location, and would like to call up information, usually in the form of a map, to help him understand where he is. This, then will be our first task – while in an unfamiliar or partially unfamiliar location, activate the system and attempt to get one’s bearings.

Route planning is a little more complicated, due to the variety of situations in which it might be used. For example, the user might be using it to assist navigation through a locale with which he is partly familiar, looking to it only when he needs clarification. This would correspond to the home user driving through an unfamiliar part of town. Or, he might use it in a completely foreign location, in a rental car for example, relying on it exclusively. These sub-tasks are worth distinguishing between as they represent different user segments and different modes of use.

Another class of tasks we consider is that of changing the route in progress. The most straightforward case would be entering a destination, following the instructions partway, then changing one’s mind, and thus the destination. The user would need to be able to figure out how to clear the old destination and enter a new one. A variation on this theme reflects a continuing difficulty with navigation systems – what does one do when the information is wrong (e.g. a road has become one-way), or when obstacles present themselves. Road work in Canada is usually performed exclusively during the summer, concentrating road closures and detours. The user must be able to deal with situations where the instructions cannot be followed for external reasons, and either compensate for it himself or somehow communicate this to the device.

Location-based services must also be considered. The intent is to allow a user to request nearby services, such as restaurants or gas stations, and have the device lead him to an appropriate establishment. The task, then, is to be able to specify such a desired service to the device, rather than a specific address.

All that remains, now, is to specify concrete objectives for our subjects to perform based on the general tasks we've identified. For the purposes of our study, we divide these into two separate sets, representing the two major groups of user task: input-oriented and direction-following directions. This will correspond to the two primary phases of our methodology, outlined in the next section.

Input-oriented tasks:

1. Subject is to use device to get a meaningful view of his current location. (Orientation)

2. Subject is to put the system into a direction-providing state, given an address. Once in a direction-providing state, subject is to change the destination address, given another.

3. Subject is to put the system into a direction-providing state for the nearest suitable service of some kind, a fast-food restaurant or gas station, for example.

Direction-following tasks:

4. Subject is to follow directions to destination in wholly unfamiliar locale.

5. Subject is to follow directions to destination in partially unfamiliar locale.

6. While following directions, subject is told that he may not drive down a particular road, as it is approached. Subject must continue to destination despite detour.

We keep it as simple as possible without losing completeness. All destinations and event locations used should be pre-planed such that the routes taken will vary sufficiently – a mixture of right and left turns, major and minor roads, and so on.

SELECTING USABILITY EVALUATION METHODS

In the software usability field, a veritable cornucopia of techniques exist for evaluating interfaces. For our study, we examined many of these usability evaluation methods, selecting aspects of those that suited our needs, constrained accordingly. First, we are interested in learnability and memorizability. Second, we are concerned with evaluating existing devices – these systems are already in the marketplace. Consequently, we avoided UEMs meant for the design stage. Finally, we are interested in looking at how “real” people fare with the devices, not the professionals behind the systems, who are unavoidably biased.

With these criteria in mind, we considered a great variety of UEMs. Methods such as cognitive and heuristic walkthrough were deemed too expert-centric for our purposes. Heuristic evaluation has already been employed by others, such as Nowakowski, Green and Tsimhoni [2], yielding valuable design insight. Other expert-centric methods were similarly rejected.

Another consideration is the distinction between methods that rely on the subject for introspection, as opposed to those that elicit information in a more indirect, less intrusive manner. For example, the method known as the question-asking protocol relies on the user's ability to answer direct questions by the experimenter as he is performing tasks. This is quite distracting and unnatural, thus unsuitable. Our goal was to yield both introspective and exhibited information without creating unduly contrived circumstances.

This further constrained our search, leading to a two-part approach - data would be obtained through 1)
surveying the users directly about their experiences, by questionnaire and interview, and 2) observation. The former is intended to elicit introspective insight from the user, the latter to pick up on what introspection cannot through observation.

The introspective component was to be achieved primarily by informal interview after the experiments, used with success in other learnability studies [3]. Users were to be asked a few simple questions regarding their experience, with follow-up questions left to the experimenter's discretion based on their responses. Users were encouraged to demonstrate using the device if possible.

For the observational component, we settled on a class of methods that employs pairs of subjects, which are particularly well suited for learnability studies. This includes the subjects-in-tandem method [5] and the teaching method. The experimenter yields insight by observing how the subjects communicate – issues that come up are dealt with verbally, but without the discomfort or bias of being queried by an experimenter. The teaching method is a variation on this theme, where the first subject familiarizes himself with the system, then instructs the other in its use.

Finally, a third method was considered, the coaching method, where the subject is given the opportunity to ask the experimenter questions regarding the device's use. Insight into the learning process and its difficulties is garnered by the questions asked. This method was deemed problematic because of the inherent problem of experimenter bias – it was feared that the experimenter might unwittingly lead the users down particular paths with respect to the device, which contradicts our purpose of understanding what users do in real situations.

Thus, we settled on a combination of the subjects in tandem method and the teaching method. This allowed us to obtain a more complete view of the user's learning process, from both the learner and teacher's perspective. Any learning issues are likely to be made apparent in at least one of the situations.

PUTTING THE PIECES TOGETHER

Having established our tasks and selected some usability evaluation methods, all that remained was to combine the pieces. Our protocol consists of 3 stages: 1) selection and pre-screening of subjects, 2) use of subjects in tandem method, focusing on input, and 3) use of teaching method, focusing on direction-following.

The first stage was implemented by way of a simple questionnaire to be filled out by potential subjects, which helped us select appropriate participants while collecting basic data on their driving experience. Since we are concerned with learnability, we required subjects with no prior experience with (electronic) navigation devices.

The survey contained a few questions to this effect, as well as some basic queries about the subject's proficiency with conventional maps and on-line map sites such as mapquest. We also inquired about how long the subjects had been driving, and briefly about their regular driving habits. The purpose of this was primarily to filter out unsuitable subjects, as well as collect basic data on those that we did use.

Information about their familiarity with various locales was used to aid in selecting locations with which they had appropriate degrees of familiarity for the direction-following tasks. We felt that it made more sense to take advantage of the diversity of our subject base by having them navigate to different locations, so long as they were suitably unfamiliar with them. More interesting issues are likely to come up in this manner. This was also much easier than finding a sufficient pool of people unfamiliar with a particular locale. We should emphasize that this study was intended to get a broad view of learnability, rather than focus on the effect of a particular variable.

Once suitable subjects were selected and vetted, they were put through the second stage. In pairs, they were put into a car with a navigation device, and the first list of tasks to perform. No user manual of any kind was provided. This stage involved input only – no driving was performed. The proceedings were recorded by a mounted video camera, so that we could examine it afterwards. The experimenter sat in the back seat to record observations and answer general questions, though he was not allowed to help the subjects figure out the device.

The experimenter was required to be very familiar with the device in question, so that he would know what to look for. Much of learnability is about recognizing when users make incorrect assumptions about how a device operates – the experimenter must know the “correct” operation of the device in order to recognize these discords.

Afterwards, the subjects were interviewed in order to record any interesting observations the subjects might make about their experience with the device. Experimenters followed up on any learning-related leads, while encouraging introspection on the part of the subjects. Questions regarding the operation and function of the device were answered, but only after all other matters. It was deemed that giving the subjects guidance in this fashion for the next stage would not be problematic, since the focus would be the learner-teacher dynamic. It was not desirable to taint their impressions from the second stage before providing whatever insight they could, however.

The third stage focused primarily on direction-following, using the second list of tasks. One participant from the previous stage was paired with a new subject. The new subject was given a set of tasks, again with locales customized to their familiarity, while the old subject was
instructed to teach the new how to use the device to this end. The old subject was not allowed to touch the device directly, so all input had to be done by the new subject/driver. The exception was any manipulation performed while driving, such as changing zoom factors, although this was discouraged for safety reasons. Once again, both subjects were interviewed for learnability-related insight afterwards. Focus this time was primarily on difficulties stemming from following the directions provided, as well as the learner-teacher experience, as it related to performing the tasks with the device.

A few more details are worth mention. First, in order to mitigate safety concerns, care was taken to choose a time of day that minimized traffic. For the first in-car phase, the experimenter’s car was used, parked in a shaded area with all equipment set up. For the second phase, the driving subject’s car was used.

RESULTS

At the time this draft is being written, the full study is still getting underway. We have, however, run a preliminary study with a small group of people to work out kinks in the methodology. Despite the small size of this preliminary run, a number of useful observations were made. The full study will be complete and full results ready in time for the final draft deadline.

The preliminary study consisted of a run with two members of our research group as subjects, followed by 4 non-practitioners, friends of the author. The latter were all between 25 and 28 years of age, with one having a technical background. Both wholly unfamiliar and partially unfamiliar routes were covered. For the full study, we will look at users between the age of 40 and 60, who are more likely to purchase higher-end cars and these systems.

The navigation device used was a Garmin Streetpilot 2620, a portable system designed for cars and boats. It was very convenient for our purposes, since it was easy to move around and set up. It is powered by the cigarette lighter via an adapter and rests solidly on the dashboard with little effort. Its interface is primarily touch-screen based, with a simple menu system and a touchscreen keyboard for address input. It also has a remote control, which was not used, and provides directions by both vocal cue and visual display.

METHODOLOGICAL OBSERVATIONS

We found the paired subjects approach lived up to our expectations. Particularly for the first set of tasks, the interaction between the subjects yielded a great deal of insight into their learning experiences, without the experience feeling unduly unnatural.

We also found that while the audio aspect of the videorecording was helpful in reviewing the experiment, the video aspect was far less so. It was difficult to effectively record the interaction with the device. Most of the interesting observations made were from the subjects’ speech. When one considers that the experimenter was also present, watching the actual interaction with the device, one can’t help but question whether full-blown videorecording is truly necessary. Thus, for the study proper, we will likely opt to do audio recording only.

ORIENTATION

Our preliminary study did reveal a significant issue to consider however relating to the orientation task - the initial state of the system. Our first experience with the device, which involved the simple task of trying to get a display of our current location and the surrounding area, proved surprisingly difficult. The zoom factor was way off (it was a continent-level display), the zooming tools were unintuitive, and it was not obvious how to centre the display about our current location, rather than arbitrary locations. When we had input our first destination, after resolving most of these issues, we were still unable to see the road in front of us despite having fixed the zoom and centering it properly. It was not until we had been driving for a little while that the system automatically adjusted the display in a sensible way.

Despite these issues, once the system had been used once, it was left in a far more usable state. The orientation task was thus trivial for the remaining users tested, because they did not have to contend with the problems that we had already dealt with. Thus, when running such an experiment it is important to consider the initial state of the system. For the full study, we will pre-set the system into a slightly less useful state, forcing the subjects to make use of the map-manipulation functionality. This will make the orientation task meaningful, as originally designed.

DESTINATION INPUT

Though more of an ergonomic issue, some users had to take the device down from the dashboard in order to manipulate it. Unfortunately this is hard for us to avoid, as some cars’ dashboards are not conveniently shaped for using these devices. For an integrated system this would not be an issue, however it is something to keep in mind for after market systems. It is also worth pointing out that this would be one reason for the inclusion of the remote control, though its functionality is limited.

Another common complaint was difficulty using the touch-screen due to having “big fingers”. Considering that none of our subjects had unusually large fingers, this could be a significant problem. It might imply that the minimal functional size of a touch-screen based system is larger than this device. We will look for this in the full study.

The non-QWERTY “keyboard” for entering addresses proved another source of complaint, even for the non-technical users. Although this might be demographic
specific due to varying degrees of familiarity with QWERTY PC keyboards, it will be interesting to see how our target demographic feels about it. It is worth pointing out, however, that research indicates QWERTY keyboards are preferable [8]. Further confusion was created by the lack of a backspace button – a delete button was present, however most users expected it to behave as a backspace button. There was also some displeasure with the manner in which the input keyboard would only display letters that lead to possible correct inputs. This was due, however to the user misspelling the address. More explicit feedback, rather than simply restricting the user from inputting certain letters, would have helped.

The problem that caused the most frustration, however, was the system's seeming knack for completely losing what the user had typed in, forcing him to re-type it. This occurred several times with all of the users, resulting in confusion (“why did it forget what I just typed?”). This seemed to be caused by the device's highly modal interface, along with some ambiguity as to how to accept or cancel. The device does not have an explicit “accept” button, expecting the user to “click” on items directly, which some users did not realize. The device's 'x' icon was particularly problematic, as it sometimes appeared to behave differently from how it would in a PC windowing system. This was the primary source of context-loss resulting in having to re-input an address. Learning transfer from other commonly used systems, then, would seem to be an issue.

Expected behaviour vs. actual behaviour is a common problem in learnability. This manifested itself in a few ways during our experiment. Addresses provided to users contained suffixes such as "street", "drive", and so on. It was not clear to them, however, if they were necessary when inputting the addresses. This caused several cases of confusion, with one user being unable to find an address that had been successfully input earlier that day by another group. This user ended up finding only a lake with the same name, after having put a space after the street name, then deciding that she did not have to put in "street" or "st". This eliminated the actual street entry, which was listed without the suffix, but was eliminated from the search results by the extra space. This lack of clarity about the device's expectations with respect to these suffixes was evident to varying degrees with every user tested. Most of them complained about it explicitly afterwards.

A similar lack of understanding about the device's expectations occurred after the address was selected. The device then provides the user with two options: "show map" or "go to". All users initially clicked "show map", under the assumption that this meant the device would provide a map showing them how to get there. When no driving instructions were provided, the users were frustrated at having to re-input the address and remember to choose the other option next time. This sometimes repeated itself. This is a clear example of a simple disconnect between the language and expectations of the designers compared to the users. "Go to" apparently did not hold the same meaning to our users. More concrete descriptions would probably solve the problem - "driving instructions", for example, is much clearer than the more abstract "Go to".

Interestingly, one user assumed the device had speech-recognition capabilities. Predicting what users expect from a device can be difficult.

Another common and more general complaint was that the device responded very slowly to input, sometimes causing the users to think it had crashed or not accepted their input (i.e. they had no pressed the touchscreen firmly enough), even though the device clearly reacted to the input by making a faint clicking noise. Long periods of blank or non-updating display were mentioned as problematic – the device would generally update at a reasonable rate, then pause for no apparent reason. This did not engender confidence in its reliability. For users, particularly technologically unsophisticated ones, to feel comfortable using the device, it must be responsive and realiable. On this note, however, it was noticed that the recalculation time when going off-route was surprisingly snappy.

Unlike the address entry tasks, the service-based task was universally accomplished with ease - no significant issues arose.

DRIVING INSTRUCTIONS

The turn instructions were generally well received, with fewer and more minor issues arising. All drivers commented that the timing of the instructions was very good, and required no explanation. Once in the direction-providing mode, users were happy with the way the device scaled the display automatically to an appropriate zoom level. Two of the three drivers mentioned that they paid little attention to the map once they realized the device would provide verbal instructions.

There were some complaints about the instructions, however. All drivers noted that the distances provided with the instructions were effectively useless – one driver expounded "why does it think I know how far 200 metres is?" upon hearing the first such instruction. She also mentioned that using an approximate time until the turn would have been far more helpful, although it did not prevent her from correctly following the instructions.

Similarly, the device initially provided directions using compass directions, which all users considered puzzling and useless. Fortunately, the device switches to turn-based directions after the car has moved a bit. It is assumed that this is a limitation of a GPS-based system, but it did cause some initial confusion.

Our distinction between navigating routes with partial familiarity and complete unfamiliarity was also validated. A few users were puzzled at the choice of route the
device picked while driving through areas they had some familiarity with - “Why is it taking us this way?”. By contrast, users seemed happy to trust the device fully while driving in unknown areas. This implies a trust issue that seem to be more significant for a user in his own locale than, say, a rental user in an unfamiliar town. This justifies our assumption that these be treated as distinct scenarios, since they yield different behaviour.

One more significant source of confusion was over lane changes and lane endings. At several points, the car ended up in a turn lane leading the wrong way. Users commented that it would have been nice for the device to warn them that they'd have to move into the left lane to continue going straight. One user also commented that he expected the device to tell him to not turn at major intersections, although he was alone in that regard. It will be interesting to see how this particular issue is borne out with a wider user sample.

One driver commented that since she was somewhat tired during the run, she was relying on the device's instructions blindly. She had been expecting the device to inform her to stay out of the right-turning lane. When it did not, she was forced to stop abruptly. This disconnect between her expectations of the device and the device's actual capabilities serves as a potent reminder of the underlying safety issues involved.

RECOMMENDATIONS

Though only preliminary results are available thus far, we have a number of useful recommendations for system builders based on them.

SIMPLIFY THE “WHERE AM I” TASK

Considering the apparent importance users place orientation [1], this task was surprisingly difficult to perform when the device was not already set into a sensible state. This should either be the default behaviour of the device upon activation with sensible zoom level, display defaults, and so on, or a very simple and obvious way to achieve the same thing be available. A single button that activates the device and displays the current location with reasonable display defaults would be ideal.

REASONABLE INITIAL STATE FOR RENTAL USERS

Rental users are often in unknown locales and thus potentially stressed or hurried. Having to expend energy trying to get a navigation device into a useable state is not likely to engender satisfaction. We recommend that for the rental market, setting the navigation device into a sensible state be part of any clean-up procedure in place when preparing a vehicle for a customer.

USE CLEAR AND UNAMBIGUOUS LANGUAGE

The “Display Map” option was consistently chosen over the “Go To” option, despite the latter being the correct on for the circumstance. A more clear label, such as “Give Directions” would have prevented a lot of hassle.

MAKE INPUT REQUIREMENTS CLEAR

Users had no way of knowing whether to input the “st” or “street” suffix when putting in the addresses given – the device did not make it clear that it expected the suffix to be omitted, despite the fact that it had the suffix information. (The full name appeared once the address was selected).

BE AS FORGIVING AS POSSIBLE

This is especially important with non-technical users, who tend to have an especially low tolerance for interface confusion. The device should have been able to accept the addresses with or without the suffix, abbreviated or unabbreviated. Users will receive addresses in many forms, and will expect their devices to accept them in whatever form they have.

PROVIDE AMPLE FEEDBACK

This is a well known but surprisingly oft ignored usability principle. The user must be constantly informed of what is happening so that they feel that they are in control of the device. A device that does not provide adequate feedback does not engender confidence and makes the end user feel helpless. This can be especially distressing when one is trying to develop a sense of how the system works for the first time. Initial impressions and trust are very important for the acceptance of new technology.

GUIDE USERS IN THE RIGHT DIRECTION – GENTLY

The input system, which restricts what letters the user can type based on possible words it can form, is useful, but initially confusing for some users. It would have made more sense to explicitly tell the user that she could not possibly form a valid address with the letters she had input thus far, rather than lead her into a corner with no explanation. A simple message would have helped her realize she was misspelling the address.

KEEP INTERFACE ELEMENTS CONSISTENT WITH OTHER WIDELY-USED SYSTEMS, IF REASONABLE

With the widespread use of PCs, PDAs, cellphones and other computing systems both at work and at home, it can be expected that many users will have some experience with electronic systems. Employing interface elements that look similar but function differently results in confusion from learning transfer. Rather than fight it, this is something that can and should be taken advantage of as much as possible. Users that can apply existing knowledge to learn the system more easily will have far less trouble with it.

FUTURE WORK
Our experience thus far has generally validated the usefulness of our evaluation method. This points to further potential studies to be done. First, in order to properly investigate the related issue of memorizability, that is how use changes over time, a longitudinal study is planned. We believe that by bringing users from this study back after varying periods of time to instruct new first-time users, we will be able to identify which elements of the interface users need to re-learn, and which elements they have trouble re-learning.

It is also worth noting that our use of the evaluation method was not focused on a specific variable – we aimed for a broad survey of users, rather than attempt to identify the effects of specific variables such as, age, availability of user manual, different types of routes (highway vs. local), and so on. By fixing parameters more tightly and varying properties of interest, a variety of more targeted studies are possible.

Finally, although the tasks used here were specific to in-car navigation devices, the general methodology is certainly applicable to other sorts of devices, in-car or otherwise. The issue of learnability has likely not been addressed adequately in some of these areas. Thus, there is no doubt opportunity to apply our methodology to examine this question elsewhere.

CONCLUSION

Our evaluation method has been shown to be effective for identifying learnability problems. The use of pairs of subjects proved non-intrusive and effective. For the device tested, address input was the primary area needing improvement, although issues were also identified for route finding. Overall, numerous problems were identified, with concrete recommendations for system builders resulting from their identification.

The fact that such issues were uncovered underscores the importance of considering learnability as a distinct aspect of the broader usability question. While existing research focusing primarily on efficiency is undeniably valuable, our learnability-focused approach complements it well, providing a more complete view of the wider problem.

Learnability should also be taken into account when developing safety guidelines and standards, particularly given the probability of infrequent and first-time use.

REFERENCES

note: The references presented below are placeholders for a proper list with full bibliographical information which will be completed for the final draft of the paper.

10. Jacob Nielsen, “Usability Engineering”

ACKNOWLEDGEMENTS

We would like to thank Auto21 and the University of Guelph for their generous support of this project.