PREPRINT. Final version available as:

BLANDFORD, A., HYDE, J. K., CURZON, P. & PAPATZANIS, G. (2008) EMU in the car: evaluating multimodal usability of a satellite navigation system. In N. Graham & P. Palanque (Eds.) *Proc. DSVIS 2008*. LNCS 5136. 1-14. DOI 10.1007/978-3-540-70569-7_1

EMU in the car: evaluating multimodal usability of a satellite navigation system

Ann Blandford[1], Paul Curzon[2], Joanne Hyde[3] & George Papatzanis[2]

[1] UCL Interaction Centre, University College London, Remax House, 31-32 Alfred Place London WC1E 7DP, U.K.
A.Blandford@ucl.ac.uk
http://www.uclic.ucl.ac.uk/annb/
[2] Queen Mary, University of London, U.K.
[3] formerly at Middlesex University U.K.

Abstract. The design and evaluation of multimodal systems has traditionally been a craft skill. There are some well established heuristics, guidelines and frameworks for assessing multimodal interactions, but no established methodologies that focus on the design of the interaction between user and system in context. In this paper, we present EMU, a systematic evaluation methodology for reasoning about the usability of an interactive system in terms of the modalities of interaction. We illustrate its application using an example of in-car navigation. EMU fills a niche in the repertoire of analytical evaluation approaches by focusing on the quality of interaction in terms of the modalities of interaction, how modalities are integrated, and where there may be interaction breakdowns due to modality clashes, synchronisation difficulties or distractions.

Keywords: usability evaluation, multimodal systems, in-car navigation systems, satellite navigation systems.

1 Introduction

There is a substantial literature on the design and use of multimodal systems, most of which takes either a system or a user perspective. Taking a system perspective, issues of concern include how to select output modalities to communicate most effectively (e.g. [9]) and how to integrate user input expressed through multiple modalities to correctly interpret the user's meaning (e.g. [15]). Conversely, much work from a user perspective is concerned with how users perceive and work with system output in different modalities (e.g. [8]) or how users select modalities of communication (e.g.

[14]). Little work has taken an integrative approach, considering both user and system perspectives in parallel. The work reported here takes such an approach, developing a prototype methodology for reasoning about the design of multimodal interactive systems to accommodate both input and output within the interaction. As an integrative approach, it does not consider the fine-grained details of either system implementation or user cognition, but focuses more broadly on how the two interact.

The method, Evaluating Multimodal Usability (EMU) was initially developed and tested using as the main case study a robotic arm interface [10]. The approach presented and illustrated here is a refinement of the method, as described below.

2 Background: multimodal interaction

Multimodal systems are widely considered to be ones that integrate multiple modes of input or output, typically using non-standard interaction devices. The standard configuration of keyboard and mouse for input and graphics, text and audio for output is rarely described as "multimodal", though for our purposes it would class as such. Many definitions of a "modality" effectively consider a data stream from a particular source. For example Lin and Imamiya [13] discuss assessing user experience by measuring various user attributes – eye gaze, pupil size, hand movements, verbal reports – and refer to each of these inputs as a modality. Similarly, Sun et al [15] discuss data fusion across speech and gesture modalities, and Oviatt et al [14] focus on how people select alternative modalities (i.e. input devices) for interacting with a computer system. Considering both user input and computer system output, Coutaz et al [5] consider the combinations of modalities in terms of Complementarity, Assignment, Redundancy and Equivalence. Here, 'assignment' means that information has to be communicated through a particular modality and 'equivalence' means that the same information can be communicated equally effectively through alternative modalities. Complementarity and redundancy refer to how information is communicated (using different modalities in complementary ways, or presenting equivalent information through multiple modalities).

From a user perspective, much work on modalities has focused on how people integrate information received through different senses. Wickens and Hollands [17] present a multiple resource theory that considers cognitive capabilities and limitations in terms of perceptual channels (vision, hearing, touch, etc.), information form (of which the two most important for our purposes are lexical and symbolic) and stages of processing. They highlight limitations on sensory input (that multiple streams of information cannot be easily received through the same channel, such as eyes or ears, simultaneously) and on input, processing and action (that competing information in the same code, verbal or spatial, cannot be easily processed simultaneously). Other approaches that take a user focus include Barnard et al's work on Interacting Cognitive Subsystems [7], which considers the transformation and integration of sensory inputs through central processing to generate action, and Kieras et al's [12] work on Executive Process – Interaction Control (EPIC), which models human information processing in complex, multimodal tasks.

Since our concern is with assessing the usability of interactive systems, the capabilities and constraints on human processing, as well as those on system processing, have to be accommodated within any definition of a modality. Drawing on the insights from earlier work on modalities, we propose a definition: that a modality is a *temporally based instance of information perceived by a particular sensory channel*. This definition comprises three key elements, time, form and channel, which need some explanation.

Time refers to the static or dynamic way in which information is presented. This element is implicit in some other definitions of modality, e.g. in work on data fusion [15], but is essential for considering how both people and systems can integrate information from multiple sources over time. We distinguish three temporal forms: continuous (meaning that information remains available, in fixed form, over an extended period of time), discrete (information is communicated once, in a transient form) and dynamic (information is communicated over time, building up the message).

Information form is particularly important from a human perspective, in terms of how people express and comprehend information. Drawing on earlier work, we distinguish three important information forms: lexical, symbolic (e.g. graphical, with a meaning that can be inferred) and concrete (e.g. a scene, where no particular interpretation is intended).

For people, the primary sensory channels are visual, acoustic and haptic, though the set of possible channels might be extended to include olfactory (e.g. [2]). If the focus were on computer system input, then alternative input channels might be considered (e.g. keyboard input typically corresponds to lexical, haptic user output).

This definition of a modality, together with the extensible classification of possible values, is at the heart of the EMU method.

3 Overview of original EMU method

The original EMU method [10] drew on earlier task-oriented structured evaluation methods such as GOMS [4, 11] to develop a process-oriented analysis of user–system interaction modalities. The central idea behind the approach was that the analyst should work systematically through an interaction, taking account of the communications between user and system and also all other environmental interactions occurring in the situation. For example, when using an in-car navigation (or sat-nav) system, both user and system are also interacting with the car and the outside world, and the driver may also be interacting with passengers, which together constitute the environment. The core question is then whether all necessary information can be received, interpreted and integrated (by both user and system).

The analysis of modalities involves considering both input and output – for both user and system. This separate analysis is necessary to consider whether all information transmitted by one of the agents (user or system) is received and correctly interpreted by the other. As well as atomic modalities, there may also be composite ones, where some modalities depend on others; for example, tone of voice may communicate information that augments or contradicts the words spoken, or the colour of a sign may convey additional information.

Particular attention is paid to modality clashes. These may be physical (e.g. a user cannot look in multiple places at once, although attention may be caught by appropriately designed visual signals in peripheral vision).They may be temporal, in that multiple information inputs may be difficult to detect or interpret if they occur at the same time: an example would be the McGurk effect, where acoustic and visual information are slightly misaligned, making interpretation of speech difficult. One particular case of clashes that affects people is that, unless trained to do so in particular situations, people are unable to process two streams of lexical information simultaneously (e.g. reading while saying something different). They may also experience semantic clashes, of which an example would be the Stroop effect, where the colour of a word clashes with the meaning (e.g. the word "blue" written in red text). Hyde [10] also recognised that some interactions might be difficult for novices, but become easier with practice, such as changing gear in a car while negotiating a bend.

In the original version of EMU, analysis proceeds through eight stages:

1. The first stage, as in many analytical evaluation methodologies, is to define the task, or tasks, to be analysed.
2. The modalities used in the interaction are then listed, both descriptively (e.g. sat-nav gives voice direction) and in terms of the modality (e.g. system expresses acoustic-lexical-dynamic).
3. The third stage is to describe the user, system and environment in terms that might have an impact on the usability of the system.
4. A preliminary assessment of any modality issues should be performed.
5. A more complete and systematic analysis is performed, listing all steps of the interaction in terms of expressive and receptive modalities, to deliver a rich account of that interaction. This may include optionality (indicated by 'or') and simultaneous communications (indicated by 'and'); it should also include a note of any preconditions for communication (e.g. that a flashing light is within the visual field).
6. Clashes, as outlined above, are explicitly considered as the next step of analysis.
7. The penultimate step is to explicitly review the modalities used, considering usability issues that emerge (e.g. over-dependence on a single modality in a situation where large volumes of information need to be communicated).
8. Finally, Hyde [10] discusses the writing of the usability report, including conclusions and recommendations.

The method has been applied to the design of several systems, including a robotic arm for use by disabled people, a ticket machine and a central heating timer.

Two substantive tests of the EMU method have been conducted: one focusing on the usability and the other on the utility of the approach.

3.1 Usability evaluation of EMU

The usability of the method was evaluated by teaching it to a group of 28 students with a background in HCI. They were asked to evaluate two systems using EMU, and their usability reports were assessed to establish how well they had understood the concepts and method, and whether they were able to apply it effectively. Following the test sessions, participants were invited to complete a questionnaire on their perceptions of applying EMU.

Details of this study are reported by Hyde [10]; here we summarise the main points of that study. The participants' detailed modality listings were compared to a model answer, and for most participants only minor errors were identified, indicating a good grasp of the core concepts. Participants experienced a little more difficulty in identifying modality clashes, and their ability to draw out usability insights from their analysis was variable. In questionnaire responses, most reported that the training had been clear, though some expressed doubts about their ability to apply the method correctly. There was, however, an overall concern that the application of the method was too time-consuming, and that an excessive attention to detail could, at times, divert attention from the broader usability issues that the analysis should have been highlighting.

The usability of any technique depends heavily on both the prior experience of the analyst and the quality of the training. As such, any single study will be inconclusive, as there are many variables that influence the outcome. Overall, the evaluation suggested that the approach could be understood and used effectively in a reasonable time, but that it should be made more lightweight where possible, without compromising the rigour of the technique.

3.2 Utility evaluation of EMU

The utility of the method was evaluated by comparing the results of an EMU analysis against those of other evaluation techniques and also against empirical data. This study focused on the design of a robotic arm for use by people with limited movement. Seven other analytical evaluation approaches were applied to the same system (GOMS, Cognitive Walkthrough, Heuristic Evaluation, PUM, CASSM, and Z and STN representations). Empirical data of the arm in use was also analysed. A full account of this study is presented by Blandford et al [1]. In brief, the analysis showed that EMU occupies a useful niche in the repertoire of evaluation techniques. Z and STN were reasonably effective at supporting the identification of system-related problems such as the lack of an 'undo' facility, redundant operators, and long action sequences. GOMS supported the identification of many of the same issues as Z and STN, plus some concerning the synchronization of user actions with system behaviour. HE identified a range of issues, as defined by the particular set of heuristics applied. Most issues identified through Cognitive Walkthrough and PUM related to possible user misconceptions when interacting with the system. CASSM covered some of the same territory as CW and PUM, but also raised issues relating to the conceptual fit between how to operate the arm controller and what the user would want to do with the arm 'in the world' (e.g. concerning how easily the user could judge arm movements). EMU also covered some of the same territory as CW and PUM; in addition, it supported the identification of various issues relating to the modalities of interaction, as outlined below.

For the robotic arm and its interface, physical considerations were important. In particular, there was scope for system misinterpretation of user intentions so that the command issued by the user was not that received by the system. For example, the gestural input device had options scrolling across the screen, and the user had to nod when the required option was displayed to select it; this depended on the user's ability to synchronise their gesture with the system state. Consistent with the motivation for developing EMU to consider multimodal issues, this method proved the strongest for identifying these issues in the interaction.

EMU also proved effective in highlighting issues concerning the dual interactions with both the arm controller (which was located within the user's natural visual field) and the arm itself (which moved within a larger space beyond the controller).

Both the usability and the utility evaluations of EMU, as well as our own experience of applying it, indicate that it is understandable and useful, but that the original version of the method was rather cumbersome to use, and therefore needed streamlining and simplifying.

4 Simplified EMU method

Following the evaluations, the EMU approach has been simplified in two ways: by reducing the number of formal stages in an EMU analysis and by de-emphasising the task analysis of the original step 5 (which typically duplicated findings of other task-oriented analysis techniques). The approach is still task-based, in that an interaction sequence, or a space of possible interaction sequences, is used as a basis for analysis, but the focus is on modalities, possible misinterpretations or breakdowns in communication, and modality clashes, considering interactions with the environment as well as with the system that is the focus of analysis.

The first stage is to select and describe a scenario of use, considering both a task sequence and the environment within which that task takes place. In order to make analysis efficient, it is important to focus attention on both representative and critical interaction sequences, but it is not necessary to consider repeated instances of the same modality configurations. When considering the environment, it is important to consider variability in the interactions; for example, in an office environment, a telephone might ring at unexpected times and distract the user, while on the road the environment provides many inputs and distractions that need to be integrated with information from the sat-nav as discussed below.

The second stage is to perform the modality analysis for selected interaction sequences. This involves considering every step, or phase, of the interaction in terms of five elements:

- *System, user* and *environment* modalities, remembering that these commonly occur in parallel (e.g. system receiving what user is expressing, or *vice versa*). In this context, we define the environment to be the broader context within which the user and system are interacting, including other technologies that are not the particular focus of the analysis.
- *Expressing* and *receiving*, noting which way the information is flowing.

- *Sensory channel*, considering for now the three possibilities of acoustic, visual and haptic, while recognising that the set of possible channels might be expanded.
- *Information form*, considering lexical, symbolic and concrete as the main forms.
- *Temporal form,* whether continuous, discrete or dynamic.

The third stage is to consider interaction difficulties. These include:

- Potential mismatches between expressed and received modalities. These might include breakdowns where information is not received at all (e.g. discrete information presented visually, but not observed, or a user talking to a computer that is not set up to receive acoustic input at that moment). They can also include mismatches due to timing or interpretation problems, such as the example discussed above of timing and the gestural input device.
- Modality clashes, as discussed above.
- Integration difficulties of making sense of information received in different modalities or from different sources. From a system point of view, this might include data fusion difficulties; from a user perspective, it may include interpretation of information in the current (environmental) context.

These stages are illustrated in the following example, where we present an outline analysis of an in-car navigation (or sat-nav) system.

5 Illustrative example: in-car navigation

In this illustrative example, we base the analysis on a Garmin system, but aim to provide a description at a level of detail that generalises across various sat-nav systems, to draw out general points, rather than to focus on the details of implementation of this particular system. As well as the details of the sat-nav design, there might also be details concerning the car, how the sat-nav is fitted in the car, what other technologies (radios, MP3 players, etc.) might be available, whether there are passengers in the car, etc. The interaction will also be influenced by details of the route travelled, how familiar the driver is with that route, what signage is available, what the visibility is, etc.

One of the challenges in defining tasks is considering the level of detail and specificity with which it should be described. This is particularly so for devices such as sat-nav systems, which are intended to respond in a rich way according to the context of use. A very detailed description might yield valuable insights, particularly if details from an empirical study are also available: this would make it possible to consider issues such as the timing of instructions, the relationship to road signs, the visibility of the up-coming junction etc. However, such a rich account might not generalise well to other (similar but non-identical) situations. Therefore, we develop an abstract description of situations to illustrate a multi-modal analysis.

5.1 Stage 1: defining the scenario: task and environment

There are usually two key phases to interacting with a sat-nav system: set-up and use. For the purposes of illustration we consider one set-up task and an abstract in-use task.

Set-up typically starts with turning the sat-nav system on, then waiting for it to start up, identify its location and present the main options. The user then has to make a sequence of selections at the interface to define a destination; the model under consideration is a touch screen device, displaying both graphical icons and text, and with acoustic output. It would be possible to critique the sequence of action steps, or the labelling of options in terms of their textual or graphical clarity, but these issues are well covered by established evaluation approaches such as Cognitive Walkthrough [16], so for this analysis we check the consistency of interaction patterns across the sequence of steps and analyse one step in more detail. The environment for set-up might be the home, car or other starting place; it is unlikely to be changing rapidly, since the sat-nav should not be programmed while driving, but it might be dark or cold, or there might be glare from sunshine.

When the sat-nav is in use, at some abstract level we have a steady state, in which the sat-nav is delivering instructions which the user is following. The external visibility may be clear or poor (e.g. glare, low visibility or dark); the road conditions may be more or less demanding (e.g. heavy traffic); and there may be distractions (such as radio) or supports (such as a passenger to interpret sat-nav information and road signs). The driver should be interacting with the car controls, but not inputting information into the sat-nav.

Fig. 1. Sat-nav menu

5.2 Set-up: modalities and possible interaction difficulties

Next, we consider steps 2 and 3 for the first task. To turn the sat-nav on, it is necessary to press and hold the 'power' button until the display lights up. At this point, the user can release the button, and wait while a start-up message is displayed (accompanied by a multi-tone bleep); the user is asked to confirm that they agree to avoid interacting with the device while driving (at which point, the user is expected to press a soft-key on the display). Every soft-key press is accompanied by both a visual rendering and an audible 'beep'. Pressing illegal options (e.g. the 'up' key when it is not possible to scroll up) results in a two-tone beep: the usual 'button-press' beep followed by a lower pitch one. These modalities are summarised in Table 1; here, the modalities columns represent stage 2 of the EMU analysis and the 'notes' column highlights possible difficulties (stage 3). For start-up, we assume that the user's attention is focused on the device, and the only likely effects of the environment are it being cold or dark, or perhaps there being glare on the screen. If the device is fixed in the car, the user may have difficulty seeing the display while interacting with it.

Table 1. Interaction modalities and potential difficulties when initialising the sat-nav

Interaction event	User modalities	System modalities	Notes
User presses power button	Expresses haptic symbolic continuous	Receives haptic symbolic continuous	Users may have difficulties locating button in the dark, or pressing the button if their hands are cold or they are wearing gloves.
System displays	Receives visual	Expresses visual	This is assurance to the

start-up message and beeps	symbolic continuous AND acoustic symbolic discrete	symbolic continuous AND acoustic symbolic discrete	user that the device is functioning (but also signifying that it is not yet ready to be used).
System displays driving warning and on-screen acceptance button	Receives visual lexical continuous AND visual symbolic continuous	Expresses visual lexical continuous AND visual symbolic continuous	The user has to recognize the 'pressability' of a soft key.
User presses soft-key to confirm acceptance	Expresses haptic symbolic discrete	Receives haptic symbolic discrete	Touch screen may not accept input if user is wearing gloves. Screen may be hard to see if there is glare or user's hand obscures display.
System displays button 'depressed' and beeps, then displays main menu. See Fig 1.	Receives visual symbolic discrete AND acoustic symbolic discrete THEN visual lexical continuous AND visual symbolic continuous	Expresses visual symbolic discrete AND acoustic symbolic discrete THEN visual lexical continuous AND visual symbolic continuous	The initial system expressions are redundant, as the user can determine the effect of their action by seeing the menu.

The initial menu is shown in Figure 1. Whereas the earlier steps can be listed succinctly in a table, this display is much richer, and each element of it can be separately analysed to assess user interpretations. Here we interleave stages 2 and 3:

- There are five areas which are represented by soft key buttons, each of which has a lexical label (with or without an additional graphical label, which might be considered redundant). These are unlikely to be problematic.
- Two other areas ('settings' and 'adjust') are also clickable, but may not be recognised as such by a novice user.
- Two other areas, the 'battery' symbol and the time, display information about the state of the device, but cannot be interacted with.
- There is an additional area that the user can interact with: the top left-hand corner, if touched, will take the user to a new display showing "GPS is Off", together with a symbolic map of local GPS transmitters and their signal strength. This is unlikely to be discovered by most users.
- There are other controls (such as the power button) and connectors around the periphery of the device. In particular, at the back of the device (not shown in Figure 1) is a hinged, but unlabelled, component: the GPS antenna. If the user raises this antenna away from the body of the device, this is interpreted by the system as an instruction to turn GPS on; then a GPS signal indicator appears in the top-left corner of the screen, and if the user touches this area then a new display shows an estimate of the current GPS accuracy. The role of the antenna as an 'on/off' switch for the GPS is not immediately apparent, and has to be learnt.

In this section, we have outlined the steps of initializing the device, and highlighted some possible usability problems that emerge from a consideration of how the user is likely to interpret system output and also how the system interprets user actions. During set-up, in which interactions with the environment (whether at home or on the road) are likely to be minimal, we have focused on the user–system interaction. In our second example, we consider the broader interaction with the environment too, and possible variations of that situation.

5.3 Driving: modalities and possible interaction difficulties

While driving, we do not consider the task structure. The set of modalities in play is relatively static, and the issues concern how those modalities interact with each other. Stage 2 involves listing the modalities:

- The system receives input from GPS satellites from which it calculates its current position.
- If there are no passengers in the car, the system is unlikely to be receiving any other inputs.
- The system expresses visual, lexical and symbolic, dynamic modalities, as illustrated in Figure 2. It also gives verbal instructions intermittently (using an acoustic, lexical, dynamic modality).
- The environment includes the external world, which in turn includes the physical context (acoustic and visual, concrete, dynamic modalities) and signage (visual, lexical and symbolic, continuous modalities).
- The environment also includes the car itself, with which the user interacts via steering wheel, foot pedals and other controls. These controls receive input from the user via touch, and the user, in turn receives feedback from the controls by haptic (and also proprioceptive) feedback.
- The environment includes other devices, such as radio and dashboard displays, in the car. Dashboard displays may take various forms, most commonly visual, lexical and symbolic, continuous. (We consider the modality of the speedometer, for example, to be continuous rather than dynamic because it changes relatively slowly, and the user does not have to monitor it continually.) In some vehicles, dashboard displays may include acoustic output (lexical or symbolic).
- Radio output (or that of other entertainment systems) is typically acoustic, dynamic, and either lexical or symbolic. If the user adjusts the radio, the modalities include haptic, symbolic, discrete modalities (for making the adjustments) and visual, symbolic (or lexical), continuous modalities (for monitoring the new system state).
- The user expresses haptic modalities in interacting with both the vehicle controls and in-car systems (e.g. the radio). In some situations, the user may also talk – e.g. if using a car phone.
- The final consideration is of user receptive modalities. The user is likely to be receiving visual (lexical, symbolic and concrete) dynamic information from the environment, visual (lexical and symbolic) and acoustic (lexical) information from the sat-nav system, acoustic (lexical and symbolic) dynamic information from in-car entertainment systems, visual (lexical and symbolic) continuous information from the dashboard, and haptic, symbolic dynamic information from the car controls.

Laying out the modalities in this way does not immediately highlight possible problems, so stage 3 involves systematically working through the modalities and identifying possible clashes and other modality problems.

First, consider the sat-nav in its environment: it has to synchronise information from multiple satellites to calculate its current position and relate that to its database of geographical information. Depending on the location and quality of signal, this may be achieved with varying degrees of accuracy, which in turn determine the quality of information the device can deliver. In some situations, poor input information can result in "Lost satellite signal", with implications for the user as discussed below. Some sat-navs also derive information from the car telemetry system, which should increase the accuracy of available information, but we do not consider this possibility further.

Fig. 2. An example sat-nav display while driving

Next we consider the sat-nav expressive modalities, and corresponding user receptive modalities. As shown in Figure 2, the device gives rich visual information, including a local route map (symbolic), the next instruction (lexical), estimated time of arrival (lexical) and distance to the next turning point (lexical). It also presents acoustic lexical information, intermittently. Systematically assessing these modalities:

- There are possible semantic clashes between acoustic and visual information from the sat-nav. The sat-nav may give an instruction such as "turn left" when the display indicates that there is still some distance to go before the turn: this information needs to be interpreted in the context of the current road situation.
- The acoustic information is, apart from the possible timing issue just noted, a subset of the visual information (in CARE terms, this is a redundant modality). If the driver only has access to the acoustic information, whether because of the location of the sat-nav in the car or because their visual attention is taken up elsewhere, the information available is relatively limited, and the meaning needs to be interpreted in the context of the external environment.
- In particular, there are possible semantic clashes between acoustic information from the sat-nav and visual information from the environment. A peculiar attribute of the auditory information from the particular sat-nav studied is that silence has a meaning – i.e. to go straight on. For example, a mini-roundabout might be regarded

by the user as a roundabout, but not be represented as such within the system, so that if the user is expected to go straight across the roundabout then the system provides no acoustic instruction (which may be particularly confusing if the direction considered to be "straight" is not immediately obvious to the user). Similarly, if the main road bends to right or left, but there is a turning that goes straight ahead, the user may be unsure whether silence from the device means that they should go straight ahead or they should follow the road round the bend. Additional information in the environment, such as road signs, may disambiguate some situations, but add further uncertainty in others. Such semantic clashes have been noted by others (e.g. [6]).

- If the user is talking, or listening to other lexical information, the dynamic nature of the acoustic instructions may result in them being missed or misheard.

Focusing just on the visual information from the sat-nav, there are possible difficulties with various elements. The thickness and colour of the line (dependent modalities) that denotes the direction indicates the importance of the route, but obscures the corresponding information about the roads to be travelled (so visual information about whether the road to be turned onto is major or minor is absent). The lexical information is changing relatively slowly, so it may be possible for the user to glance at these information items while also attending to the road, but their relatively small size may make glancing difficult (depending on the location of the sat-nav in the car). The overlaying of some information (e.g. the white arrow over the "M25" label in Figure 2) and the placing of lexical items (e.g. "ters Crouch" in Figure 2) means that there is uninformative visual data on the screen. The display shown in Figure 2 is only one of several alternative displays: we use it for illustration purposes rather than evaluating all possible information presentation forms on this device.

Turning attention to the acoustic information alone, we note that it has various attributes: content (e.g. "turn left in 0.3 miles"), timing and tone. It has already been recognized (e.g. [3]) that the timing of instructions relative to the external environment is critical as the verbal information from the sat-nav needs to be interpreted in the context of the physical situation: for example, there may be ambiguity over which turning to take when there are several in quick succession. In EMU, the tone of voice is considered a dependent modality: for the sat-nav, it always sounds equally confident and reassuring. This may be at odds with actual degree of certainty (e.g. due to poor satellite information, or inaccuracy in underlying data), and consequently mislead the driver. We surmise that many incidents of people following sat-nav directions while ignoring warning signs in the environment are at least partly accounted for by the authoritative tone of voice employed in most sat-nav systems.

As noted above, most acoustic information is a replication of visual information. There are some exceptions, notably information about the state of the device. One example is implicit information that the driver has failed to follow directions – indicated by the acoustic information "recalculating". This contrasts with the information that a navigating passenger would typically provide, which would continue to refer to the navigation problem (e.g. telling the driver they have just gone the wrong way). Another example is "lost satellite signal", which leaves the user unsure how to interpret any subsequent directions, or leaves the user in an unknown place, with insufficient information to make decisions about where to go at junctions.

This second example illustrates some of the abstract issues that can be identified in a static analysis, focusing on the available modalities, the information that they communicate, and possible clashes and breakdowns in that communication.

6 Conclusions

Multimodal and ubiquitous systems are becoming widespread. Established analytical evaluation techniques are not well adapted to identifying the usability issues raised by the use of alternative or multiple modalities, or of assessing how systems are used within their broader environments. EMU is an approach that can complement more traditional evaluation techniques by focusing attention on information flows around an interactive system within its broader environment which will, itself, typically transmit and receive information that may augment, complement or interfere with that which passes between user and device.

We have illustrated the application of EMU to identifying some of the limitations of an exemplar sat-nav system. A satellite navigation system was chosen for this study because: the interaction between user and system is multimodal; the use of the system only makes sense within the broader environmental context (of the geographical region being traversed); and the system is safety-critical, making usability and user experience particularly important.

EMU focuses attention on the modalities of communications between user and system within the context of use. In particular, clashes between modalities, integration of information, and possible lost information can be identified through an EMU analysis. Earlier studies have shown that EMU is learnable [10] and that it occupies a particular niche within the space of analytical evaluation methods [1]. However, the initial study of EMU in use highlighted the fact that the process of conducting an analysis was unduly laborious. In this paper, we have presented a more lightweight approach, such that the costs of analysis are more appropriate to the benefits gained through conducting that analysis.

Acknowledgements

The development of the original EMU method was funded by a studentship from Middlesex University for Joanne Hyde. Recent work has been funded by EPSRC grants GR/S67494 and GR/S67500.

References

1. Blandford, A., Hyde, J. K., Green, T. R. G. & Connell, I. (forthcoming) Scoping Usability Evaluation Methods: A Case Study. To appear in *Human Computer Interaction Journal*.
2. Brewster, S., McGookin, D. & Miller, C. (2006) Olfoto: designing a smell-based interaction. In *Proc. CHI '06*. ACM, New York, NY, 653-662.

3. Burnett, G.E. (2000) Usable vehicle navigation systems: Are we there yet?, Vehicle Electronic Systems 2000 - European conference and exhibition, ERA Technology Ltd, 29-30 June 2000, pp. 3.1.1-3.1.11
4. Card, S. K., Moran, T. P. & Newell, A. (1983) *The Psychology of Human Computer Interaction*, Hillsdale NJ: Lawrence Erlbaum
5. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J. & Young. R. (1995) Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In K. Nordby, P. Helmersen, D J Gilmore, and S Arnesen (Eds.): *Human-Computer Interaction: Interact'95*. Chapman and Hall, 1995. pp. 115-120.
6. Curzon, P., Blandford, A., Butterworth, R. & Bhogal, R. (2002) Interaction Design Issues for Car Navigation Systems. In Sharp, Chalk, LePeuple & Rosbottom (Eds.) *Proc. HCI 2002 (Vol. 2).* (short paper) 38-41. BCS
7. Duke, D. J., Barnard, P. J., Duce, D. A. & May, J. (1998) Syndetic Modelling. *Human-Computer Interaction*. 13, 337-394
8. Elting, C., Zwickel, J., and Malaka, R. 2002. Device-dependant modality selection for user-interfaces: an empirical study. In *Proc. IUI '02*. ACM, New York, NY, 55-62
9. Fink, J. & Kobsa, A. (1998) Adaptable and Adaptive Information Provision for All Users, Including Disabled and Elderly People. *New Review of Hypermedia and Multimedia*. 4. 163-188.
10. Hyde, J. K. (2001) *Multi-Modal Usability Evaluation*. PhD thesis. Middlesex University.
11. John, B. & Kieras, D. E. (1996) Using GOMS for user interface design and evaluation: which technique? *ACM ToCHI* 3.4. 287-319
12. Kieras, D. E., Wood, S. D. & Meyer, D. E. (1997) Predictive Engineering Models Based on the EPIC Architecture for a Multimodal High-Performance Human-Computer Interaction Task. *ACM Trans. Computer–Human Interaction*. 4, 230-275.
13. Lin, T. & Imamiya, A. 2006. Evaluating usability based on multimodal information: an empirical study. In Proceedings of the 8th international Conference on Multimodal interfaces (Banff, Alberta, Canada, November 02 - 04, 2006). ICMI '06. ACM, New York, NY, 364-371.
14. Oviatt, S., Coulston, R. & Lunsford, R. (2004) When do we interact multimodally? Cognitive load and multimodal communication patterns. *Proc. ACM International Conference on Multimodal Interfaces (ICMI)*. 129-136.
15. Sun, Y., Chen, F., Shi, Y., & Chung, V. (2006) A novel method for multi-sensory data fusion in multimodal human computer interaction. In *Proc. OZCHI '06*, vol. 206. ACM, New York, NY, 401-404.
16. Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994) The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. Mack (eds.), *Usability inspection methods*. New York: John Wiley. 105-140.
17. Wickens, C. D. & Hollands, J. G. (2000) *Engineering Psychology and Human Performance*. Prentice Hall International, London (Third Edition).