Distinctive Aspects of Mobile Interaction and their Implications for the Design of Multimodal Interfaces

Luca Chittaro

HCI Lab
Dept. of Math and Computer Science
University of Udine
via delle Scienze, 206
33100 Udine, Italy

http://hcilab.uniud.it

ABSTRACT. People want to do more with their mobile phones, but their desire is frustrated by two classes of limitations. One is related to the device, its hardware and software. The other is related to the context, and comprises perceptual, motor, cognitive and social aspects. This paper will discuss some of the opportunities and challenges that this complex scenario presents to multimodality, which can be a key factor for a better design of mobile interfaces to help people do more on their mobile phones, requiring less time and attention.

Keywords: mobile devices, multimodal interfaces, attention, cognitive workload mitigation, human factors, human-computer interaction.

Introduction

People want to do more with their mobile devices, but their desire is frustrated by different kinds of limitations. A pioneering work on studying the peculiar aspects of interaction in mobile conditions was carried out by Pascoe et al. [22]. They focused on a specific class of users (fieldworkers, in particular archaeologists and ethologists) and on a specific task (fieldwork data collection) pointing out that mobile devices could effectively support it, provided that the following requirements are taken into account: (i) allow to collect data whenever and wherever the fieldworker wants, whether she is standing, crawling or walking, (ii) minimize the amount of user attention required to interact with the device, (iii) enter high volumes of data quickly and accurately, (iv) have the device sense the environment to record some context information such as location, and help in analyzing it (e.g., plotting her observations to a map, based on the locations in which they have been recorded).

In this paper, we aim at advancing the characterization of interaction in mobile conditions and its implications for design, dealing with several issues that have not been touched by [22]. In particular, we (i) extend the analysis to mobile users and mobile tasks in general, (ii) extend the set of distinctive aspects of mobile interaction that are taken into account, (iii) extend the interface design space to multimodal interfaces. First, we will concentrate on the differences between interaction with mobile and with desktop system. Special consideration will be given to attention and distraction issues as well as the need for user's cognitive workload and stress mitigation. Then, we will illustrate the consequences on the design of mobile multimodal interfaces, and provide some general recommendations on how to approach it.

Distinctive aspects of mobile interaction

This section presents the peculiar aspects of mobile interaction that distinguishes it from interaction with desktop systems, making it more difficult to build effective user interfaces for mobile users.

Differences between interaction with mobile and desktop systems

First, mobile devices are limited in their input-output capabilities (screen, keyboard, buttons, sound,…). For example, while the latest years have seen significant increases in the size and resolution of the displays we use in the office or at home, allowing us to organize our tasks better and carry out more complex tasks with less effort, nothing similar has happened on mobile phones: mobile displays remain small and the size gap with respect to desktop displays is constantly widening. Besides peripherals, less powerful hardware (CPU, memory, buses,...), limited software support for multimedia and multimodal applications, and slower connectivity contribute to make it harder to develop powerful interaction techniques on mobile devices. It must be stressed that several of these device limitations are not likely to disappear in the near future because mobile devices need to remain compact in size and energy-efficient.

Second, the context has much deeper effects on interaction when we consider users on the move, instead of users operating devices (mobile or not) while sitting at a desk in their office or home. These effects show at different levels:

- Perceptual: unlike office or home, the physical parameters (illumination, noise, temperature and humidity, vibration and motion,…) of the mobile user's environment are extremely variable, limiting or excluding one or more modalities. For example, in a noisy street we can become unable to perceive sounds from the mobile device; under a glaring sun, we can be unable to discriminate color on the screen or even to read the screen at all, on a moving vehicle we might not notice vibrations generated by the device,…).
- Motor: mobile conditions can impair user's ability to fine control her voluntary movements or to take specific postures (e.g., standing). For example, the accelerations and decelerations of a vehicle subject passengers to involuntary movements, which interfere with motor operation of the device (e.g., time and errors in selecting options, effectiveness of using writing recognition or gesture recognition software,…).
- Social: even when using certain modalities would be perfectly possible from a perceptual and motor point of view, social norms related to different environments may make it impossible or unadvisable. For example, keeping sound on at a conference is not tolerated, while looking at the device screen is accepted; keeping a lit display in the darkness of a movie theatre is not tolerated; making wide gestures while being near strangers might be embarrassing in general and not advisable in specific places (e.g., hospitals, police stations,…) or when the strangers are oversensitive to aggressive cues (e.g., a group of hooligans).
- Cognitive: people in mobility conditions can devote only a very limited attention to interacting with applications on the device. Unlike office and home environments, when we are in a street or at the airport or in a bar, we have to attend to a constant flow of events and stimuli that come from the environment, and respond to those events with proper actions. Some of these events can affect the success of our travel (e.g., hearing on the airport public address system that our flight has been moved to another gate), others our social relations (e.g., listening and properly responding to what people who

are with us are saying), others even our personal safety (e.g., noticing potential dangers while we are in a street). Although many of the events happening around us will be uninfluential, we have to devote them some attention anyway to decide that they are so. This makes using the mobile device a secondary rather than a primary task, and significantly limits the amount of cognitive resources that the user can allocate to interacting with the device. It can also contribute to increasing users' level of stress.

The issues highlighted above at the perceptual, motor, and social levels could be faced by offering a wide repertoire of modalities (and combinations of modalities) to the mobile user, allowing her to choose those that are socially accepted, easier to perceive and easier to operate in the context she finds herself in (for example, the user might turn sound off at a conference and have sound output translated into visual or haptic stimuli, she might choose to give commands to the interface through a touch screen or keyboard while waiting at the hospital but later give the same commands through gestures when she finds herself in a more private environment, and so on). It must also be mentioned that a wide repertoire of modalities would also turn helpful in making mobile phones more accessible to disabled users (for example, Ghiani et al. [11] built a PDA-based mobile museum guide for blind users, exploiting two vibrotactile actuators and speech output, for moving along a path of tagged objects).

The issues highlighted above at the cognitive level are instead more insidious, and require to rethink mobile interaction in terms of cognitive workload mitigation. To do so, research is needed in two major directions. First, we have to understand better how human attention operates while using devices in mobile conditions. Second, we have to devise proper mobile interface design strategies for cognitive workload and stress mitigation. The following two sections are devoted to discuss each of the two directions.

Attention and distraction issues in mobile interaction

The more attention an interface requires, the more difficult it will be for the mobile user to maintain awareness of the surrounding environment and respond properly to external events, risking the negative consequences we have seen in the previous section. A considerable amount of research has been carried out on a specific mobile context: using mobile phones, navigation systems or other on-board interfaces (e.g., radio and CD player, Internet services) while driving a car. This research has proved that distraction caused by mobile devices is a real safety danger as well as characterized different types of distraction and studied some of the effects of different modalities on distraction. Although some results are necessarily specific to the driving task, more general findings – such as the distinction between general and selective distraction and the effect of different modalities on them – apply more generally to the various contexts of mobile interaction.

Green [12] summarizes some of the major conclusions reached about the role played by distraction caused by mobile devices in inducing car accidents. For example, the risk of a collision when using a cellular phone is up to four times higher than when a cellular phone is not used, and units that allow hands-free interaction seem to offer no safety advantage over handheld units. Moreover, reaction times while using the phone increase up to 60 percent, and glance data suggests decreases in attention to the road due to using the phone while driving.

In an experiment carried out by Hancock et al. [13], 42 licensed drivers were required to respond to an in-vehicle phone while faced with making a crucial stop decision at an intersection equipped with a traffic light on a closed-course test track. In this dual task, divided attention condition, there was a slower response to the stop-light presentation and drivers compensated this delay by subsequently braking more intensely. Even worse, an increase in non-response to the stop-light was recorded, which corresponds to increased stop-light violations on the open road. Safety risks are exacerbated by the fact that many drivers are unaware of their decreased performance while using mobile phones, as pointed out by Lesch and Hancock [17] who collected subjective ratings of performance from drivers who had to deal with phone-related distracting tasks while driving.

Three different types of user's distraction caused by mobile devices and on-board interfaces have been identified. Distraction manifests itself mainly as withdrawal of attention, which can be of two types: *general* and *selective*, according to the classification by [1], or *eyes-off-the-road* and *mind-off-the-road* according to the classification by [12]. A general withdrawal of attention (eyes-off-the-road distraction) is due to eyelid closure (in the case of fatigue) or eye glances away from the road (in the case of visual inattention). Tasks that are visually demanding, such as reading detailed maps or long strings of text, lead to a greater amount of eyes-off-the-road distraction [12].

Selective withdrawal of attention (mind-off-the-road distraction) seems to be a more insidious type of distraction. In this case, although the driver's eyes look at the road, driver's attention is focused on thought (e.g., listening to a long or complex auditory message). This can lead to a selective filtering of information based on expectations rather than the actual situation and to *looked-but-did-not-see* phenomena, which can have serious consequences: for example, a driver at a cross-road looks at a vehicle approaching from the right without really seeing it, and crosses the road colliding with the other vehicle.

The third form of distraction is *mechanical interference*. This refers to body shifts out of the driver's neutral position, e.g., reaching a mobile phone or leaning to see or manipulate a device, and can degrade the driver's ability to execute manoeuvres. It must be noted that using hands-free or eyes-free devices contributes to respectively decrease mechanical and visual distraction, but does not imply improvements on cognitive distraction.

In contrast to the large number of studies on driver's interaction with mobile devices, surprisingly little has been done with pedestrians. An interesting work in this direction has been carried out by Oulasvirta et al. [21]. A mobile phone was instrumented with 2 lightweight minicameras, one pointed towards the phone (display and keyboard), the other towards the user. Two additional minicameras were respectively placed on the user's shoulder facing forward, and on the experimenter who followed the user to record the overall environment. Users carried out Web search tasks on the mobile phone in different settings: one was a laboratory setting, the others were different kinds of mobile settings (busy street, quiet street, metro platform, cafeteria, railway station, escalator, bus, metro car). Analysis of the recorded materials showed that while user's attention devoted to the device had spans of over 16 seconds in the laboratory setting, only bursts of just a few seconds were achieved in difficult mobile situations, the number of attention switches was 8 times larger in the mobile context, and carrying out the Web search task in certain situations took away resources devoted to actions in the surrounding environment such as the control of walking, leading researchers to describe users' attention in mobility as "impulsive, fragmented, and drastically short term". In a related paper [24], Roto and Oulasvirta suggest that multimodal feedback can be particularly useful to prevent the user from looking at the mobile phone screen when she does not need to. In particular, for Web browsing, they recommend to use non-visual feedback when loading a Web page requires more than 4 seconds. This way, the user is not required to perform periodic visual checks of the screen to notice if the loading of the page is complete because she can get the information through auditory or haptic feedback.

Cognitive workload and stress mitigation in mobile interaction

In human-computer interaction, strategies for cognitive workload and stress mitigation have been traditionally studied in the design of safety-critical interfaces (e.g., aviation, nuclear power, chemical plants, weapons systems,…). The work done in these domains has been surveyed by Szalma and Hancock [26] who organize the different strategies in two major classes: changing the user and changing the task. In the first class, cognitive workload and stress are mitigated by selecting persons who prove to have better performance in the task or by training users to make them more skilled and resilient in facing the task under stress. In the second class, the interface of the system is designed to impose a smaller cognitive workload on the user, and two major directions are identified: (i) display design: under stress, complex problem-solving and analytical skills are the most vulnerable and decline first, so providing displays of information that can be perceived directly with a minimum of information processing requirements can be very effective, (ii) adaptive automation: when the user is overtaxed, another way of unloading her is to have the system assume control of some task functions. However, in the contexts considered by this paper, these traditional strategies are only partially valid. In particular, strategies aimed at changing the user are hardly viable. First, mobile devices are meant to be used by the general public, and not limited to users who have the best physical and cognitive abilities or substantial background knowledge. Second, although training could be in principle considered (e.g., through software applications pre-installed on the device itself), the need for considerable training to become able to use the mobile device is unlikely to make it popular among consumers. Most of the mobile design effort should thus go towards making users' tasks less demanding. From this point of view, display design and adaptive automation can both be exploited. However, with respect to display design, we have to take into account the previously described limitations of mobile devices, which offer less opportunities than an airplane cockpit or a plant control room. With respect to adaptive automation, we could instead take advantage of the increasing availability of sensors on mobile devices to exploit context-awareness for adaptive automation purposes (e.g., location sensing is already exploited by commercial applications for purposes that range from automatically panning and zooming maps based on user's position and speed to automatically tailoring user's queries to Web search engines in such a way that the geographically closest results become more relevant).

Implications on the design of Mobile Multimodal Interfaces

While the previous sections have already analysed some of the effects of the mobile context on multimodal interaction, in this section we expand the topic with a deeper discussion of eyes-free interfaces and with indications about how to approach the design of a mobile multimodal interface.

Eyes-free interaction and its limits

The mobile interaction literature is giving increasing emphasis to "eyes-free" interfaces, and appealing applications have been proposed, ranging from speech-based information readers to new uses of non-speech sound and vibrotactile stimuli in mobile contexts. An example of how traditional mobile widgets can be redesigned to allow for "eyes-free" use is provided by the earPod [29], a circular touchpad which can be divided into up to 12 sectors: a user wearing headphones can hear the name of the menu item located under her finger when she touches a sector, and sector crossing produces a click sound. If the user lifts the finger when it is over a menu item, that item is selected and a "camera-shutter" sound is produced; if she lifts the finger when it is in the center of the touchpad, no item is selected.

Other researchers have made mobile interfaces “eyes-free” without resorting to speech. For example, the Shoogle [28] translates some basic information about messages (number, length, type of sender) received on a mobile phone into non-speech sound and haptic information. To get the information, the user has to shake the phone, thinking of it as if it were a box containing balls. Each message is metaphorically mapped into a ball and the effects of shaking the imaginary box are simulated to produce the corresponding audio and haptic feedback. For example, hearing that one of the balls is distinctly metallic, and has a deep, heavy feeling impact, the user realizes that a long text message from a colleague has been received. The sensation of a number of lighter, glassy objects indicates instead several short messages from various friends.

Eyes-free interfaces may also benefit from richer proprioceptive feedback in touch interaction. For example, Murray-Smith et al. [19] exploit a device that provides a range of textures in the surface of its case, allowing the user to stroke, rub, scratch or tap. In particular, they implemented an interface for a music player, which allows volume and track choice to be controlled by scratch-based interaction while keeping the device in a pocket.

In general, although the “eyes-free” concept can be effective and appropriate for simple tasks as those mentioned in the examples above, uncritically adopting it as a priori specification in the design of mobile multimodal interfaces is not necessarily beneficial. It is important to consider carefully what the actual scope of eyes-free interfaces could be under the light of the cognitive perspective and the variability and complexity of contexts pointed out by this paper. We should in particular take into account that:

- Sight is the dominant sense for most users, allowing them to receive more information and more quickly through the eyes rather than other senses.
- Using external visualizations and exploiting visual processing abilities in carrying out a complex, information-rich task is a typical (and successful) strategy humans use to decrease cognitive load and simplify the task.
- From an attention and cognitive load perspective, making an interface “eyes-free” simply shifts human acquisition of information from one sense to another and does not necessarily guarantee that the user will be able to better handle the task. More specifically, moving the acquisition of information from

the visual to the audio and/or haptic channel guarantees that the user will be able to continuously look at the surrounding environment but does not guarantee that her performance will improve with respect to an interface that requires sporadic glances to a display. For example, consider the distinction between eyes-off-the-road and mind-off-the-road distraction in the previous section: an eyes-free interface undoubtedly reduces the first, but how much does it increase the second? In other words, an eyes-free interface might be more attention-demanding and distracting than the non eyes-free interface it is meant to substitute.

- Although being currently moving in the physical environment is a strong motivation for resorting to eyes-free interfaces also from the safety point of view, it would be wrong to assume that the eyes-free interface eliminates safety risks. For example, with a series of closed-course driving experiments conducted with 41 drivers ranging in age from 19 to 70, Cooper et al. [9] showed a clear negative effect of audio messages on driver performance.
- Aesthetics is a crucial factor in orienting consumer choices concerning mobile devices, and visually pleasing, graphical interfaces play an important role also from this point of view (see e.g. the iPhone interface and applications). Moreover, interfaces of desktop products increasingly resort to graphics, so users can partly exploit their training and familiarity with those visualizations when they move to mobile devices.

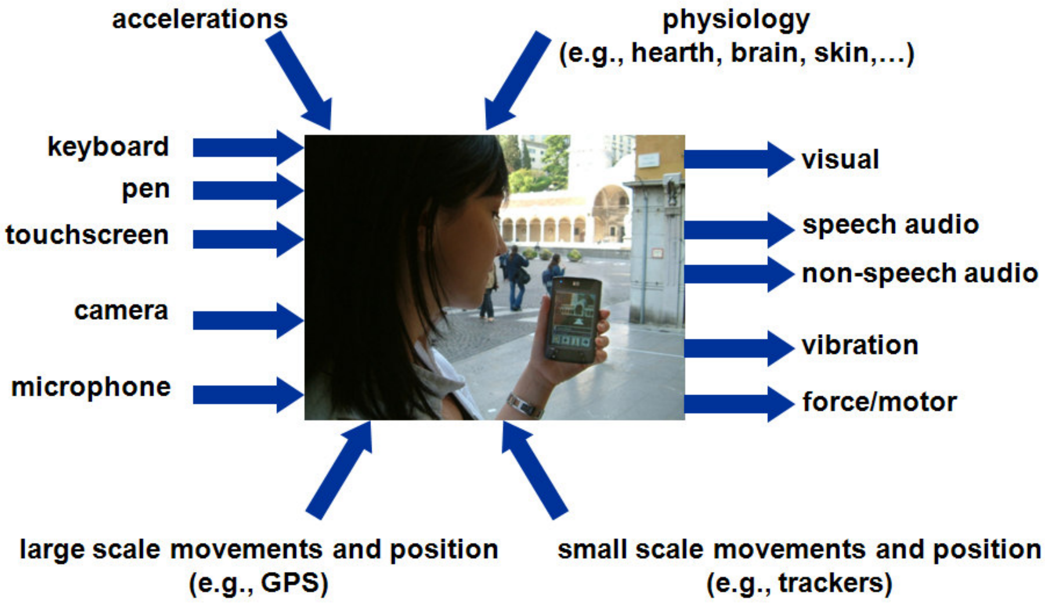

Figure 1. Input and output channels in mobile multimodal interfaces.

Designing Multimodal Interfaces for the mobile context

Based on the considerations made in the previous sections, design and evaluation of mobile multimodal interfaces should give priority to user's attention, performance, stress and cognitive workload issues, without a priori biases against or in favor of specific senses. Figure 1 summarizes the many channels that can be currently exploited to send and receive information from a mobile device. The general goal is to design for minimal attention (or, to put it another way, "permanent distraction"), not just move the same information to another channel. More specifically, in designing the mobile interface we should:

- Send to each chosen sense information in the form and amount which can be quickly and easily processed by that sense. Similarly, when information is not acquired automatically by the device but has to be provided by the user, the interface should receive it in the form and amount which the mobile user is easily able to handle in minimal attention conditions.
- Choose appropriate combinations of modalities. Multimodal feedback or input does not result in enhanced performance per se. Indeed, although each modality in isolation could have been well designed and effective, some combinations of modalities can be incongruent, not synergic and can even lead to interference (e.g., receiving speech messages concerning something unrelated to what we are currently reading on the screen is not easy to process). It is thus not sufficient to combine two (or more) modalities that work well in isolation, but one has to test the effect on users of the combination, even for simple widgets. For example, different combinations of visual and haptic feedback in terms of perceived congruence to implement a button widget on a mobile screen have been studied in [15] .
- Exploit context awareness to minimize attention requirements and cognitive workload. This involves several levels:
    - Information, the device should present the user only with the information relevant to properly handle the task at hand.
    - Modalities, the device should choose the best modality or combination of modalities based on the task and context to suit different mobility situations.

- Functions, the device should offer the functions which could be useful or wanted by the user in the current situation, but also consider if the cognitive workload required by those functions can be safely handled by the user. Regardless of the “anytime, anywhere” slogan frequently associated to mobile products, we should accept the fact that not any task should be carried out at any time and place. Cognitive workload considerations are starting to show in commercial products, e.g. navigation systems which disable manual destination entry while the vehicle is in motion because the task would be too demanding for the user or prevent the on-board phone to ring while the user is engaged in complex driving activities (inferred from data such as speed, acceleration, pedal positions,…).

From this perspective, to develop detailed guidelines for designing mobile multimodal interfaces, research is needed to determine which modalities and which combinations of modalities are most effective in the different perceptual, motor, social, and cognitive conditions a mobile user can find herself while she carries out given tasks. For example, Perakakis and Potamianos [23] studied form filling on PDAs using a typical flight reservation task and identified multimodal synergies between speech and visual interaction, which when combined gave better results than traditional GUI interaction alone and eyes-free speech-only interaction (the latter was the less efficient way of filling the reservation forms). Cox et al. [10] concentrated instead on the SMS input task and compared the common multi-tap and predictive text input techniques of mobile phones with speech input. The speech-only interface supported higher input speeds than traditional multi-tap and predictive text interfaces, but combining modalities improved the results: in particular, using keypresses for navigation and speech for text-entry gave the best speed. Cherubini et al. [8] considered the task of tagging and retrieving pictures taken with a mobile phone, carrying out a user study that compared three different input modalities: typed text only, speech only, or both. Interestingly, the users’ subjective preferences they obtained are inconsistent with those of the above mentioned lab study conducted by [23]. While in the lab study users showed a preference for speech input over text, in the study by [8] users carried out the task also in the field and privacy played a greater role: participants felt sometimes uncomfortable speaking to their phones in public places. On the other hand, consistently with the other studies mentioned above, entering long tags required more time with typed text. Based on these results, Cherubini et al. recommend to support both modalities so that the user can choose the most convenient based on the perceived level of privacy and availability of time. Moreover, since users who had the opportunity to use both modalities did not remember which they used for tagging specific photographs, the authors of the study recommend automatic conversion of audio input into text and vice versa to support crossmodal retrieval.

Hoggan et al. [16] have instead studied at what exact levels of environmental sound pressure and vibration, audio and tactile feedback become respectively ineffective in a touchscreen typing task. Having this knowledge also opens the possibility to make multimodal interfaces adaptive, allowing them to automatically choose the channels which are more appropriate to the current mobile context.

Considering cognitive workload of mobile users has implications on evaluation of user interfaces too. Ideally, user testing should be conducted in the intended mobility conditions. When that is not possible, the experimental procedure could artificially simulate similar cognitive workload conditions. This is often achieved through dual-task assignment. For example, to test a mobile haptic and graphical interface in the lab, Leung et al. [18] had subjects perform a task on the interface while listening in parallel to audio letter sequences played to them waiting for a specific subsequence to be recognized.

Case study: supporting users in navigating the physical world

A particularly relevant case study for mobile multimodal interfaces is given by applications which support pedestrians in navigating the physical world. These applications are affected by all the distinctive aspects of mobile interaction discussed in this paper, including safety issues. Given the fact that the less the pedestrian has to look at the screen, the more she will be able to keep her eyes on the environment, some projects have focused on developing eyes-free navigation support. For example, Holland et al. [14] have attempted to build an eyes-free interface to GPS that sonifies information about distance (using metaphors such as "Geiger counter" or "warm/cold") and direction (using panned sounds across the stereo sound range), but carried out only an informal user test that did not compare user's performance obtained with the sonified GPS vs. traditional GPS. Other researchers proposed haptic approaches, e.g. Van Erp et al. [27] indicate direction and distance through vibrations, using a belt of 8 actuators placed around the user's waist to cover the cardinal and oblique directions. Also in this case, a user study on pedestrians has been conducted but was focused on comparing different haptic distance-coding schemes, and not differences in performance between haptic and traditional navigation systems.

In our work [6], we started from the approach followed by current commercial navigation systems, i.e. combining visual information on the display with speech audio directions, and explored if it could be improved through a better display design to require less attention and reasoning time on the user's side. The solution we proposed was based on displaying location-aware photographs of the surrounding environment with superimposed perspective arrows (see Figure 2). Experimental evaluation of this approach compared to the traditional GPS map-based approach showed a significant improvement in the time required to a sample of 12 tourists to reach destinations in a city. In particular, a detailed analysis of the improvement revealed that it was due to a significant decrease in the amount of time (less than half) users spent when they had to stop to orient themselves by looking at the information on the screen.

In general, it would be interesting to evaluate combinations of the different modalities mentioned above (speech and non-speech audio, haptic, minimal-attention graphics) to test which combinations lead to enhanced performance in following directions from a navigation system.

However, it should be noted that navigating an environment often requires carrying out activities (such as planning) for which the support of a map - internal (cognitive) or external (on the mobile device) - is essential. The map should be able to provide an overview of the environment (for high level identification of different areas and targets) as well as detailed views (for lower level identification of landmarks, street names, intersections,…). Handling this variety and amount of information without resorting to visual information would be hard for the interface designer and the user as well. We thus focused specifically on exploring how interaction with maps on mobile devices can be made more effective. Interfaces that enhance spatial memory are particularly important for the mobile user, since they would reduce user's need to frequently re-orient herself in exploring the map and re-evaluate the location of the intended targets and points of interest. In [2], we proposed ZEN, a touchscreen-based solution to explore maps on mobile devices that visually provides information about the proportions and location of the currently displayed portion of map with respect to the whole map, and centralizes usually separated zoom and pan controls into a single widget. Contrasted with the two map exploration techniques most frequently provided by commercial applications, ZEN produced significantly better results in terms of user's spatial memory performance. In [3], we also introduced the possibility of visually specifying queries on geo-referenced data and visualizing their results on the map, to better help the user in planning and taking decisions with maps on the mobile device.

Besides providing navigation information through multiple modalities, a further challenge is to go beyond the traditional limitation to location in the context-awareness capabilities of navigation systems and exploit all the different channels summarized in Figure 1. For example, in [4], we have integrated a pulseoxymeter into a mobile navigation system so that we are able to provide audio instructions also about the proper speed to maintain, based on cardiac fitness considerations.

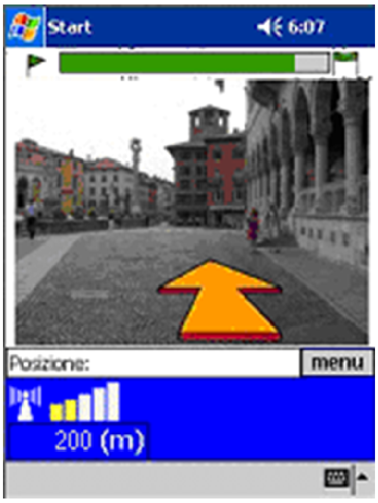

Figure 2. Providing visual navigation instructions using photographs and arrows [6].

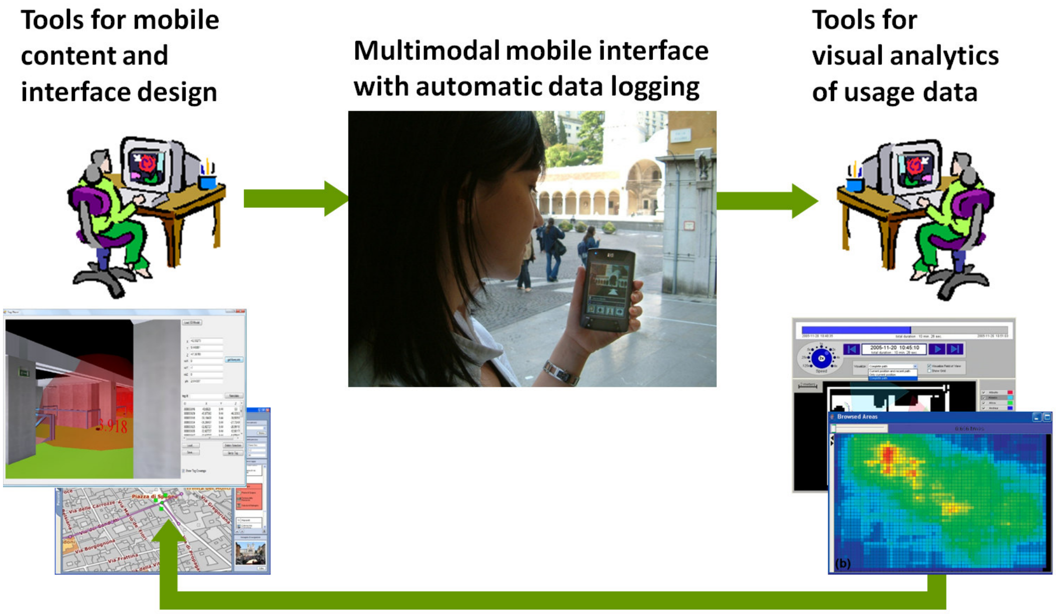

Figure 3. Iterative development scenario for multimodal mobile interfaces.

Final Remarks

This paper has analysed some of the many differences existing between mobile and desktop interaction, pointing out how they impact design and evaluation of mobile multimodal interfaces, and stressing the need for research about which modalities (and combination of modalities) are effective for different tasks in mobile conditions.

In closing, it must also be mentioned that engineering multimodal mobile interfaces would benefit from new tools that could help to understand better how mobile users exploit or respond to different modalities. For example, analysis tools of usage data could automatically log all user data exchanged through the various modalities and then present the designer with informative visualizations of that data at different levels of detail. Tools of such kind that we developed in our work allow us to log data such as user interface actions on the touchscreen of the mobile device [5], position of the user in the environment [7], physiological parameters [20], and then study all this data on a desktop or laptop system using detailed (e.g., VCR-like replay) or abstract (e.g., heat maps) visual analytics techniques.

Another useful class of tools should concentrate on rapid prototyping of mobile multimodal interfaces, to support an easy exploration of different modalities and their combinations as well as on-the-fly tuning of the parameters of each modality. For example, the OI Framework [25] aims at achieving this goal by supporting an open source repository of reusable and adaptable components, that the designer can quickly assemble together in different ways. Tools for rapid prototyping of multimodal mobile interfaces would allow the designer to make changes to the interface in the field without requiring much time or having to go back and forth from the lab as it often happens today. From this perspective, these tools should also consider content besides interaction, e.g. making quick in-the-field changes to the database of points of interests or the maps of a location-based application.

Figure 3 summarizes and depicts the scenario that would be supported by the availability of the above mentioned tools. Designers would rapidly prototype the mobile multimodal interface and content on a desktop system (or a laptop in the field), the interface support environment would automatically log all data exchanged through the various modalities during user tests, and the results of the analysis of this data would feed back into the design of the content and the interface, effectively supporting an iterative user-centered design approach.

Acknowledgements

The author acknowledges the financial support of the Italian Ministry of Education, University and Research (MIUR) within the FIRB project number RBIN04M8S8.

References

1. Brown I. D. Driver fatigue. Human Factors, Vol. 36, No. 2 (1994) 298-314
2. Burigat S., Chittaro L., Gabrielli S., Navigation Techniques for Small-screen Devices: an Evaluation on Maps and Web pages, International Journal of Human-Computer Studies, Vol. 66, No. 2 (2008) 78-97
3. Burigat S., Chittaro L., Interactive Visual Analysis of Geographic Data on Mobile Devices based on Dynamic Queries, Journal of Visual Languages and Computing, Vol. 19, No. 1 (2008) 99-122

4. Buttussi F., Chittaro L., MOPET: A Context-Aware and User-Adaptive Wearable System for Fitness Training, Artificial Intelligence In Medicine Journal, Vol. 42, No. 2 (2008) 153-163
5. Burigat S., Chittaro L., Ieronutti L., Mobrex: Visualizing Users' Mobile Browsing Behaviors, IEEE Computer Graphics and Applications, Vol. 28, No. 1 (2008) 24-32
6. Chittaro L., Burigat. S. Augmenting Audio Messages with Visual Directions in Mobile Guides: an Evaluation of Three Approaches, Proc. 7th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI), ACM Press, New York (2005) 107-114
7. Chittaro L., Ranon R., Ieronutti L., VU-Flow: A Visualization Tool for Analyzing Navigation in Virtual Environments, IEEE Transactions on Visualization and Computer Graphics, Vol. 12, No. 6 (2006) 1475-1485
8. Cherubini M., Anguera X., Oliver N., de Oliveoira R., Text versus Speech: A comparison of Tagging Input Modalities for Camera Phones, Proc. 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI), ACM Press, New York (2009) 1-10
9. Cooper P. J., Zheng Y., Richard C., Vavrik J., Heinrichs B., and Siegmund G. The impact of hands-free message reception/response on driving task performance. Accident Analysis and Prevention 35 (2003) 23–35
10. Cox A.L., Cairns P.A., Walton A., Lee S. Tlk or txt? Using voice input for SMS composition. Personal and Ubiquitous Computing 12 (2008) 567-588
11. Ghiani G., Leporini B., Paternò F. Vibrotactile Feedback as an Orientation Aid for Blind Users of Mobile Guides, Proc. 10th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI), ACM Press, New York (2008) 431-434
12. Green P. Crashes Induced by Driver Information Systems and What Can Be Done to Reduce Them (SAE paper 2000-01-C008). Convergence 2000 Conference Proceedings, Society of Automotive Engineers, Warrendale, PA (2000) 26-36
13. Hancock P.A., Lesch M., and Simmons L. The distraction effects of phone use during a crucial driving maneuver. Accident Analysis and Prevention 35 (2003) 501–514

14. Holland, Morse, Gedenryd, AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface, Personal and Ubiquitous Computing, vol. 6, no. 4 (2002) 253-259
15. Hoggan E., Kaaresoja T., Laitinen P., Brewster S. Crossmodal Congruence: The Look, Feel and Sound of Touchscreen Widgets, Proc. International Conference on Multimodal Interfaces, ACM Press, New York (2008) 157-164
16. Hoggan E., Crossan A., Brewster S., Kaaresoja T. Audio or Tactile Feedback: Which Modality When? Proc. 27th international conference on Human factors in computing systems, ACM Press (2009) 2253-2256
17. Lesch M. F., and Hancock P. A. (2004). Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? Accident Analysis and Prevention 36 (2004) 471–480
18. Leung R., Maclean K., Bertelsen M.B., Saubhasik M.. Evaluation of haptically augmented touchscreen GUI elements under cognitive load, Proc. International Conference on Multimodal Interfaces, ACM Press (2007) 374-381
19. Murray-Smith R., Williamson J., Hughes S., Quaade T. Stane: Synthesized Surfaces for Tactile Input, Proc. 26th SIGCHI conference on Human factors in computing systems, ACM Press, New York (2008) 1299-1302
20. Nadalutti, D., Chittaro, L. Visual Analysis of Users' Performance Data in Fitness Activities, Computers & Graphics, Special Issue on Visual Analytics, Vol. 31, No. 3 (2007) 429-439
21. Oulasvirta A., Tamminen S., Roto V., Kuorelahti J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI, Proc. 23th SIGCHI conference on Human factors in computing systems, ACM Press, New York (2005) 919-928
22. Pascoe J., Ryan N., Morse D. Using While Moving: HCI Issues in Fieldwork Environments, ACM Transactions on Computer-Human Interaction, Vol. 7, No. 3 (2000) 417–437
23. Perakakis M., Potamianos A. A Study in Efficiency and Modality Usage in Multimodal Form Filling Systems. IEEE Transactions on Audio, Speech & Language Processing 16 (2004) 1194-1206

24. Roto V., Oulasvirta A., Need for non-visual feedback with long response times in mobile HCI, Proc. 14th international conference on the World Wide Web (WWW), ACM Press, New York (2005) 775-781
25. Serrano M., Ramsay A., Nigay L., Murray-Smith R., Lawson J. L., Denef S. The OpenInterface Framework: A tool for multimodal interaction. Proc. 26th SIGCHI conference on Human factors in computing systems, ACM Press, New York (2008) 3501-3506
26. Szalma J.L., and Hancock P.A. Task loading and stress in human-computer interaction: theoretical frameworks and mitigation strategies. In Human Computer Interaction Handbook, 2nd Edition, Lawrence Erlbaum Associates, USA (2008) 115-132
27. Van Erp J.B.F., Van Veen H.A.H.C., Jansen C., Dobbins T. Waypoint Navigation with a Vibrotactile Waist Belt, ACM Transactions on Applied Perception, Vol. 2, No. 2 (2005) 106–117
28. Williamson J., Murray-Smith R., Hughes S. Shoogle: Multimodal Excitatory Interfaces on Mobile Devices, Proc. 25th SIGCHI conference on Human factors in computing systems, ACM Press, New York (2007) 121-124
29. Zhao S., Dragicevic P., Chignell M., Balakrishnan R., Baudisch P. Earpod: eyes-free menu selection using touch input and reactive audio feedback, Proc. 25th SIGCHI conference on Human factors in computing systems, ACM Press, New York (2007)1395-1404