Empowering people rather than connecting them

Roderick Murray-Smith
Department of Computing Science,
University of Glasgow
Scotland, UK

Abstract

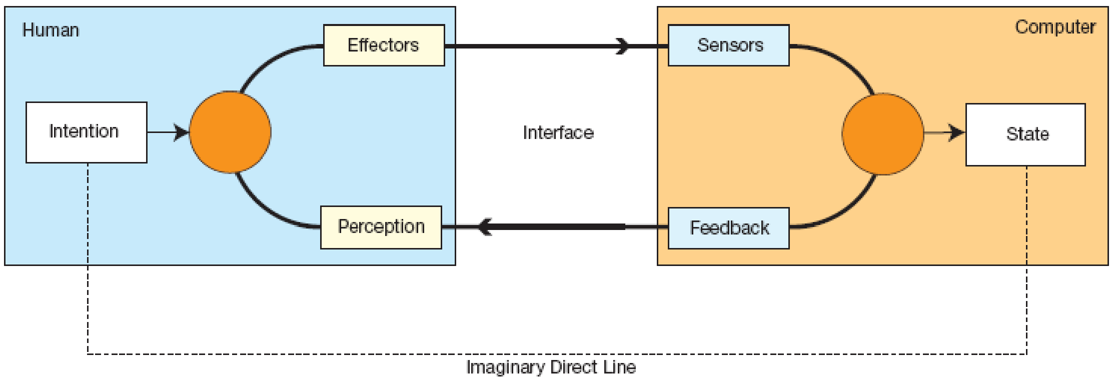
This paper discusses the consequences for the fundamentals of interaction design given the introduction of mobile devices with increased sensing capability. Location-aware systems are discussed as one example of the possibilities. The paper provides eight challenges to the mobile HCI research community, and makes suggestions for how the International Journal of Mobile HCI could contribute to the field.

Introduction

The history of mobile interaction has been largely about creating smaller and smaller devices with more and more computing power, faster connections and better displays. Now, increasingly, mobile phones are being better equipped with more sensing capability via physical sensors, and they are accessing further information from the Internet via sources which can be viewed as 'virtual sensors'. These developments give users the chance to create embodied interaction loops in a range of novel ways. In this paper I try to summarise my perspective on the fundamentals of interaction design and the challenges we face, and which I believe we need to address with our research and development effort in the coming years.

Interaction as closed-loop design

In this paper, one of the basic assumptions about the fundamentals of interaction is that a user's behaviour is about them controlling their perceptions (Powers, 1973). We view the closed loop between user and phone as a dynamic system, where designers can alter the feedback mechanisms in the phone, and where, to an extent, the human user can adapt, in order to create and appropriate closed-loop behaviour.



There is already a wide variety of sensing and display technologies that can be used to construct the physical aspects of a human-computer interface, and much recent research has been dedicated to expanding the sensing and display capabilities of typical devices. Rich sensors, from accelerometers, smart clothing and GPS units, to pressure sensors, create the potential for whole new ways of interacting with computational devices in a range of contexts. Each of these has different information capacities, noise properties, delays, power demands, frequency responses, and other modality-specific characteristics. Sensors will continue to get cheaper and smaller, and new ones will create as yet unimagined interaction possibilities. Building interfaces that make use of possibly high-dimensional, noisy, intermittently available senses to create usable communication media is a significant challenge for the current HCI framework. We need general frameworks which are not tied to specific sensing or display devices, but generalise to wider classes of devices.

The display in any human-controlled control system is to provide the user with information needed to exercise control, i.e. to predict the consequences of control alternatives, to evaluate the current status and plan future control actions, or better understand the consequences of recent actions. Current examples of basic feedback loops include: Visual, audio, or vibrotactile display of the states of phone, or of distant events, people or systems.

In a mobile context users are subject to significant levels of disturbances and tend to have a lower attention span, leading to fragmentary or *intermittent* interaction. Because of this, in many cases it can be advisable to use *modality scheduling*, where the order of presentation of information in different feedback channels can be controlled as a function of the context, and the user's control behaviour. Perception is commonly seen as process of receiving information from the world, which is typically followed by cognitive processes and then action. However, in reality, perception is tangled up with specific possibilities of action, so perception is not mediated by action-neutral models. Inner states are 'action-centred'. Gibson called this detecting 'affordances'. Such affordances are nothing other than the possibility for use, interpretation and action offered by the local environment to a specific type of embodied agent.

Perception is not passive reception of information – it is geared to tracking possibilities for action. Traditional actions on a mobile phone consisted of button pushing, but in a modern phone the action might also include gesturing with the phone, tapping the phone, walking or driving to a new location, changing the phone's compass bearing. Some of these are essentially discrete actions, but many others are actions which are spread out over time, and which may be difficult to classify consistently. This perspective changes the design of feedback for users, the software engineering platforms devices are developed on, and has implications for the users' models of how to engage with their mobile devices.

It also changes the notion of the sort of services users can be offered. One can imagine 'Apps stores' offering customers new sensory perception abilities (e.g. a 'mother-in-law early warning system' which provides feedback when the mother-in-law is within 4 minutes of your current location, or a 'my friends are in the pub' sensory input, which is activated when more than 2 of your close social network contacts are co-located in a pub), which could then be combined with actuation modules which generated particular behaviours, such as changing your visibility status, when initiated. Frameworks which make it easy for end users to combine multiple sensor readings, and possibly integrating them with other information from the Internet, allows the creation of 'virtual sensors' which offer a new service. These can then be shared with other users, many of whom will have little technical skill or interest, but who can profit from the work of others.

This approach can be summed up as *empower the users – don't dictate*! Give users more options, and create opportunities to build control loops combining different groups of output and input channels. Let the users create meaningful interactions, and let groups of users put information and structure into the world, creating the conditions for their future interactions. As discussed at length in (Clark 2008), humans are the ultimate examples of beings which are suited to create and adapt to new ecological niches. There are also some interesting objective measures of empowerment as the maximum information flow the agent can direct into its future sensoric input via the environment. It is a measure of control suggested by (Klyubin, Polani & Nehaniv 2006), building on work of Powers (1973), and is the information-theoretic capacity of an agent's actuation channel. As Klyubin *et al* say, an agent which maximises empowerment, is continuously striving for more options, which lead to more potential for control or influence as if to the motto *"All else being equal – keep your options open".*

Once people's behaviour starts to be shaped by the properties of these novel sensory-action loops we can expect to see interesting novel, but hard to predict developments in the way people move and behave in our cities. In this issue, Pirhonen and Sillence note the effect of the mobile phone use on the behaviour of young Finnish boys. The flexibility provided by the mobile phone allowed them to defer decisions on what to do next, and led to less frequent games of ice hockey. Once people add extra sensory perceptions to their mobile devices which are a function of their behaviour, location and proximity to others in their social network, we will see interesting and unpredictable new developments in such situations. Maybe it becomes enough for three boys to start heading towards the ice hockey for others to be alerted to this and join them?

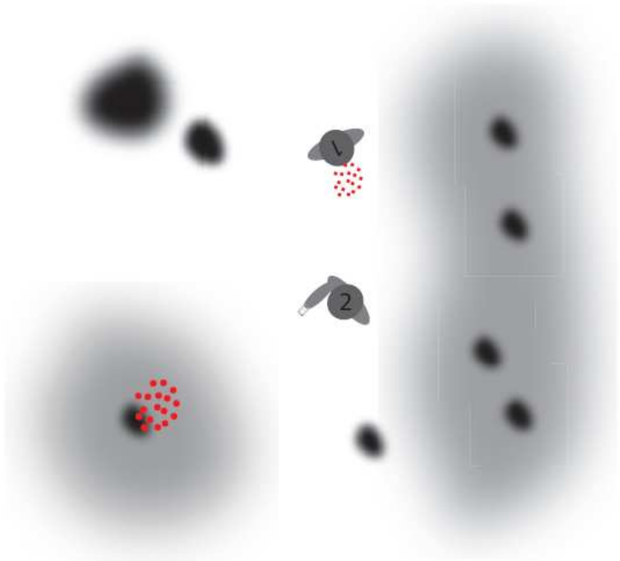
We will see an increasing level of conscious and unconscious offloading of information into the environment, which might be for our own use, or for others'. This is *stigmergy* – indirect communication between agents via the environment (which in our case includes the digital environment) without targeting a specific recipient. The academic work in this field is used to explain nest building and foraging in social insects, but it is also a vital step in the creation of cultural instruments such as language. Maybe the three boys had not actually originally intended to play another game of ice hockey, but as they formed a core of a potential game, others might be attracted to that area, and a game would emerge in a self-organised manner.

We might see a more general move in interaction design away from our simplistic 'command-and-control' systems towards a metaphor which is more like dancing, with a natural ebb and flow of control between the user, their mobile device and the broader environment. In such *negotiated interaction* systems, the control changes fluidly as the context determines. The user and the computational system work together to negotiate their interactions and communicate their intentions in a fluid, dynamic manner (Williamson, 2006). Timed, informative feedback shares the load between both sides, and the interactions occur at multiple time-scales. We are currently developing such prototypes in a collaboration between Glasgow and Swansea Universities in a joint research project in Negotiated Interaction.

This change of perspective makes us consider whether we should maybe view the interaction more as we do with animals? Here we would think of a rider on a horse, rather than someone giving a human butler instructions. The rider ‘reads’ the horse, and the horse reads the rider’s intentions via body language, gait, general behaviour, pulling on reins etc. The human and horse each have something to bring into the interaction, and are aware of different constraints, and there will be a varying level of control, depending on these constraints. In their paper on the ‘H-metaphor’ Flemisch *et al.* (2003) discussed the loose/tight reined approach to cooperative control in detail, in the context of cockpit automation. This metaphor seems very appropriate for the world of mobile interaction, where increased sensory interaction can provide the device a sense of context, but where this is subject to significant uncertainty.

Location-based interaction

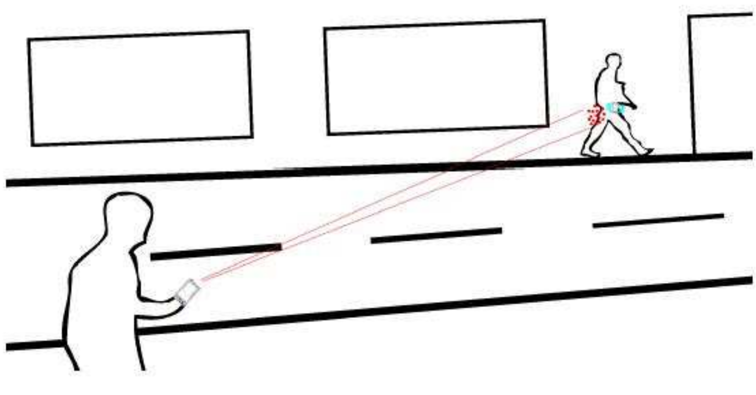
A clear example of the potential benefits of offloading information into the environment is that of location-based interaction, where content and services are placed in the environment for the user to access as needed, in a context-dependent fashion. While there is already a lot of work in this area (Fröhlich et al 2007), Mobile Spatial Interaction (MSI) feels very much like where the world-wide web was in 1992, when the first browsers came out. The huge potential can be seen, but there is still a long way to go, and we need to couple location with an ability to use bearing and projection information (Strachan & Murray-Smith, 2009). Without the ability to point from your current location in multiple directions, and at different distances, the location-based service options remain fairly limited



User 1 is ‘the cursor’ while User 2 has the ability to point at different locations by changing their bearing and the distance they are pointing.

A user interacting with digital content and services in his physical garden

As an example of where this technology might take us, we include an illustrative mobile use-case scenario: On the left of his garden footpath the user has recorded the location of a number of work-related virtual objects and on the right, leisure objects. When he leaves for work in the morning he wants to pick up e-mail for that day and any new voicemails. Stepping out of the house, he queries the area where he has stored his work objects by pointing at them. His location is detected using GPS, his bearing via a magnetometer. He feels, via tactile feedback, that he is pointing at the 'work' area of his garden, and he probes around this context, finds the email and gauges the amount of new mail by force-feedback, and grabs them from that location with a gesture, feeling a satisfying clunk as they transfer to his device. He then pulls in an upbeat music recipe from the leisure context on the right and heads for his train. Approaching the station, he points in the direction of his train for an update on any travel delays, with a pulsing vibration indicating the time to the next train. At the end of the day he comes home again and decides to leave his work behind him, so he locates again his work context by the footpath and drops his email there, effectively redirecting all work-related emails until the next morning when he can pick them up again. He then probes the leisure area, detecting its specific haptic feedback, feels his usual news podcast there but decides to leave it, and instead takes a relaxing playlist for the evening ahead, goes inside, and grabs a beer.



Geosocial networking will bring embodied interaction into the world of social networks

How will MSI cultures evolve? Will it be a top-down 'designed' approach, with major infrastructure investment by content providers and operators, or will it be a bottom-up approach where users create local content and services which are of value to them personally, and where they share this with others, shaping the evolution of their behaviours and the content they interact with in space and time? We would, as with other mobile technologies, expect to see variations in evolution, given same technology but varying local cultural and physical constraints. Will cities sponsor local content to guide people to specific locations? This could be the standard tourism scenario, but it might also be used by the police, where they could try to shape behaviour in the real world by sponsoring content and service availability which would attract young people away from areas where there had been disruptive behaviour, towards areas where the police felt things were more controllable? Will retailers give away content with gradients designed to attract people closer to their 'bricks and mortar' shops? Will operators provide location-specific functionality?

Challenges

The preceding sections have outlined some of the broad issues in the evolution of novel forms of interaction in the world of sensor enriched mobile devices. This section is an attempt to highlight some of the interesting research challenges we need to solve in order to make progress.

1. *The Midas touch.* How do we control the interpretation of our phone's sensor readings? How do we 'declutch' certain modes? The information flow from the sensors will need be interpreted differently in different contexts. This requires excellent models to automatically infer likely intention given overt behaviour, and we need subtle feedback to the user for them to infer current mode and consequences of action. This is a major, fundamental area which will recur everywhere in mobile multimodal interaction.

2. *Inference, learning and adaptation.* Understanding the data generated by mobile users: To survive in mobile service provision, companies will need the skills to create models, infer latent variables, such as context and goals, and be able to negotiate these interpretations with the users. Interpretation of context, of music content, of emotional aspects of behaviour and the meaning of interactions within social networks.

3. *Uncertainty* Every interaction involves uncertainty. An interface should be honest - it must work with the uncertainty and not just filter it out blindly, as is the norm today. Increasing computational power makes this *ad hoc* approach less defendable. The quality of control of a system depends on its feedback. The feedback must reflect the uncertainty of system beliefs. Appropriate use of uncertain feedback can lead to smoother interaction, with user behaviour regularised appropriately (e.g. Körding & Wolpert, 2004).

4. *Physical simulation for interaction design* Physical simulation based models for real-time synthesis for audio and haptic rendering. The model for this could be the development of the tools within the computer-generated film industry, where complex physical models are now controlled by artists with little need for them to understand the underlying physical models. In mobile software engineering, we therefore need to be enabling interaction designers to create dynamics, the way animators create films. This is likely to be done via standard parameterised primitives, using a range of physical metaphors, with accompanying vibration and audio feedback profiles. The quality of output is open-ended - the models can

progressively improve as computational power and models improve, and as better sensing and displays evolve. This will then lead to the evolution of richer cultures of linking such models to abstract data structures.

5. *Real –time mobile Operating Systems* For a sense of flowing, embodied control, we will need mobile devices with real-time architectures, with hard real-time guarantees. This will involve tight loops from sensing to actuation on device (including fast vibrotactile and audio responses, which are not subject to random delays due to multi-thread OS, as in current popular platforms). It will also need guaranteed low-latency round-trip-times over the wireless network for interaction between people at a distance. This would lead to major rethink of mobile operating systems requirements.
6. *Power implications of control & inference loops* This issue combines aspects of the 'midas touch' challenge, and the real-time mobile challenge, with hardware and software design and the choice of interaction metaphors. Without appropriate hardware and algorithms for intelligent allocation of limited processing power, none of these techniques will be used in everyday mobile devices.
7. *Aesthetic-utilitarian trade-off* Will interactions be so rewarding people enjoy working hard to 'master' them? E.g. In Mobile Spatial interaction, can we provide feedback gradients as elegant as a Zen garden? Will well-designed multimodal interaction make the device feel as rich as a fine musical instrument.
8. *Dancing or commanding?* Can we create metaphors which users can accept which explicitly support interaction which consists of a smooth ebb and flow of control between human and machine, rather than a dialogue of instructions and responses? Your phone can actively engage with you at a range of conceptual levels and timescales. On a scale from autistic to spookily mindreading, how alive to your wishes do you want your phone feel?

Outlook for the International Journal of Mobile HCI:

How could the International Journal of Mobile HCI contribute to progress towards the challenges outlined above?

I think we need theoretical frameworks which can handle the sensing, modelling and inference which will be such a vital aspect of modern mobile interaction, and an openness for interaction design which breaks the 'discrete-event' mould.

We will also need to make sensor interpretation and signal processing developments which enable truly engaging context aware interaction, and we should respect and publish the research in this area, rather than viewing it as the 'technical engineering details' which are hidden at the back of a user study. Without appropriate sensing and intelligent power management, none of the futuristic goals can be realised in practice. Many of the most interesting applications of the inference technology might end up being used at very low levels, in order to sense and respond rapidly to perceived changes in intention.

Furthermore, when we work with such highly instrumented devices, we should use the sensors not just for interaction design, but also for instrumented usability analysis, when we do our scientific studies. Standard mobile phones now have sensing and storage abilities which allow us to log significant amounts of data, including location, activity levels, behaviour, and voice characteristics during normal use, and to use this to better understand the effects of design decisions on user behaviour, especially if multiple members of a social network are instrumented. This allows users to go into more realistic environments, but we sense more of the activities and at a finer grained level than before. And we can use automated behaviour or context detection to initiate experimental interventions, regaining some level of experimental control, with the ecologically appropriate conditions (Roto *et al*. 2004, Oulasvirta *et al.* 2005). Researchers will be faced with demands for increased levels of objective data and more explicit behavioural metrics to augment and complement the more traditional subjective feedback, and the research community needs to explore, calibrate and standardize the use of such metrics.

References

A. Clark, Supersizing the mind: Embodiment, Action, and Cognitive Extension, Oxford, 2008.
O. Flemisch, A. Adams, S. R. Conway, K. H. Goodrich, M. T. Palmer, P. C. Schutte, The H-Metaphor as a Guideline for Vehicle Automation and Interaction, NASA/TM—2003-212672, 2003
Fröhlich, P., Simon, R., and Baillie, L., Mobile Spatial Interaction. Personal and Ubiquitous Computing, Vol 13, No. 4, April, 2009.
R. J. Jagacinski and J. M. Flach. Control theory for humans : quantitative approaches to modelling performance. L. Erlbaum Associates, Mahwah, N.J., 2003.
K. Körding and D. Wolpert: Bayesian integration in sensorimotor learning. Nature, 427:244–247, 2004.
A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: A Universal Agent-Centric Measure of Control. In Proceedings of the 2005 IEEE Congress on Evolutionary Computation, Vol. 1, p128-135. IEEE Press, Sep 2005.
Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J., Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. Proceedings of CHI 2005, ACM Press, New York, pp. 919-928.
W. T. Powers, Behavior: The Control of Perception, Aldine de Gruyter, 1973.
Roto, V. Oulasvirta, A. Haikarainen, T. Lehmuskallio, H., & Nyyssönen, T., Examining mobile phone use in the wild with quasi-experimentation. HIIT Technical Report 2004-1, August 13, 2004.
S. Strachan, R. Murray-Smith, Bearing-based selection in mobile spatial interaction, Personal and Ubiquitous Computing, Vol. 13, No. 4, 2009.
J. Williamson. Continuous Uncertain Interaction. PhD thesis, Dept. Computing Science, University of Glasgow, 2006.

Acknowledgments

We are grateful for support from: SFI grants 00/PI.1/C067 and 00/RFP06/CMS052, the EPSRC project EP/E042740/1, the IST Programme of the European Commission, under PASCAL Network of Excellence, IST 2002-506778, and OpenInterface Project. Nokia provided a donation of funds and equipment which further supported this work.