Context and Situation as Enablers for Multi-Modal Interaction in Mobile Games

Holger Mügge
University of Bonn
Bonn, Germany
muegge@cs.uni-bonn.de

Pascal Bihler
University of Bonn
Bonn, Germany
bihler@cs.uni-bonn.de

Mark Schmatz
University of Bonn
Bonn, Germany
schmatz@cs.uni-bonn.de

Armin B. Cremers
University of Bonn
Bonn, Germany
abc@cs.uni-bonn.de

ABSTRACT

Location-based mobile multiplayer games can offer a unique and intensive game experience. Three spheres of experience merge: social interaction, virtual game world, and the real location. While the potential is huge the main obstacle to a seamless and enjoyable experience is the user interaction with the mobile system. The problem is not the sheer lack of interaction techniques, but the assignment of appropriate modality in different game situations. The main statement of this paper is: multimodality alone is not enough to cope with mobile gaming. Context-awareness that lead to situation-specific interaction modes may come to the rescue. This paper discusses the requirements of modality selection against the background of location-based multiplayer games and illustrates the correlation to context-awareness and situation detection.

1. MARKET

In the last years, mobile phones moved from simple voice- and text-handsets to small computing platforms. With this, the possibilities for enhanced entertainment options have increased: photos, video, music – everything on the go. Since the advent of the Java-based phones a vivid and growing mobile games market has evolved. A recent study by Juniper research[1] expects the mobile games market to reach 10 billion US-$ by 2009 and more than 460 million mobile users are expected to download games for their devices in that year. When the iPhone AppStore launched on July 10, 2008, more than 160 out of 500 applications provided gaming experiences.

Currently most of the discussed mobile games focus on a *casual* single player play mode, but powerful location-aware mobile devices enable more games that exploit location and other context as well as social interaction to create games with rich experience. This market share has recently shown the biggest growth. Hence, the development of such games will be an issue of growing importance in the next years.

[1] juniperresearch.com/shop/viewpressrelease.php?pr=63

HCI September 1-5, 2008 Liverpool, Great Britain

2. GAMES IN MOTION

Location-based multiplayer games use the location of players, but the way they use motion is different. Explorative games or location-based puzzles can have a rather casual attitude towards location. In this paper we focus on those games where location and movements are of major importance. To characterize this family of games we give a definition of motion intensive games here:

> **Motion intensive games** *depend* on the movements of the players, taking place in the *physical reality* and influence the game in *real time.*

Typical exponents of motion intensive games are rallies, hide-and-seek games, etc. They are characterized by continuously moving players, frequently interacting with the virtual game world and communicating with teammates or opponents. We currently develop *Scotland Yard to go*, an adoption of the classical board game *Scotland Yard*[2] where a group of detectives works together to catch Mr. X in the city of London. In the original board game the focus is on strategic pondering about different theories regarding Mr. X's current location which is only revealed every 5th round. Movements are round-based and only symbolic in terms of relocating tokens on a board. In contrast, *Scotland Yard to go* is played in the real, supported by wirelessly networked mobile computers.[3]

So, what specific requirements do these games impose on HCI? They often feature situations which request time-critical game-interaction. In such situations the game activities require the player's full attention and his capabilities to interact with the device are very limited. As an example, during a hide-and-seek game the fleeing player can not interact with a computer by typing while running 'for his life'.

[2] Developed in 1983 by Ravensburger Spieleverlag, English edition published in 1985 by Milton Bradley/Hasbro

[3] The current version of *Scotland Yard to go* demands UMPCs running Windows XP. The next version will use Apple's iPhone.

Even observing a small display while moving seems unfeasible (more detailed examples are presented in section 5). The lack of attention the user can afford for the interaction with the computing device has to be considered when designing an intuitive controllable motion intensive game.

3. SENSING IN MOTION

Moving needs *vision*, hence while we move our vision is highly stressed. Consequently visual interaction with a display is limited or nearly impossible when running. Semi-transparent headmounted displays or retina projectors are a potential solution as they can display information overlaying the real field of vision. Unfortunately, they are not widely available soon so we can not rely on them. Therefore, game design for motion intensive games should not rely solely on visual interaction in situations where motion is crucial.

Audition is not as much stressed by movements as vision. Although in some situation audible capacity is reduced too (e. g. while talking face to face with teammates, or due to background or traffic noise), it can take the pressure off the players visual attention. Even though information transfer is limited by a narrow bandwidth and pure sequentiality, acoustic signals are valuable as they can be noticed in parallel to other audible perception and thus are relative independent of the player's situation. Emotional influence by sounds and music is a specificity of the media that can be used to enhance the game experience and communicate the course of the game on a high and rather vague level. Brewster gives in [2] an overview on using audio for non-speech output. There are plenty of studies attesting the influence of music on perception and performance that could be exploited for motion intensive as well as dramatists games.

Tactition can also be used for interaction. A vibrating device for example can communicate signals in parallel to other communication, while the player is moving. There are some special effectors like vibration belts or tactile interacting shoes that transfer more complex information like the direction, but they are far from being standard equipment. Nevertheless, the standard mobile device contains a simple vibration effector that can be modulated in frequency, sometimes even in intensity. We envision to apply vibration frequency to communicate vicinity. Hemmerl et al. already used it to emit a continuous 'sign of life' to communicate the overall status of the mobile phone. Nevertheless, vibration effectors can only transfer very few information and their perception is unreliable.

In addition to tangible output, contact differentiation also plays a major role in input interaction. A perceptable difference for different input modes (as known from the 'F' and 'J' keys on a standard PC keyboard which support touch typing by small raisings) can help to interact with a mobile device in situtions where vision controlled input is not applicable. Tactons (tactile icons) are a concept for tactile displays developed by Brewster, Brown and others, c. f. [3]. Even systematic deformation of tangible devices can be used to increase the information transfer. First prototypes of dynamic knobs have been developed (c. f. [8]) but they are too far from maturity to consider them here in more detail.

Other human senses can currently not be used in practise. Smell and taste can not be reached by computing devices. *Thermoception* could theoretically be used but there are no effectors available on the market today[4]. Practically it already works, e.g. when you notice the warmth of your computer and get alert that it did not switch off correctly. But you can use it only in very special situations, e.g. in a crime thriller game as a clue. *Nociception* has already been used by computer game designers[5] but cannot be stimulated with standard mobile handsets. *Equilibrioception* might be stimulated by certain *optical effects*, but no studies which apply this on a mobile device are known to the authors. We discuss a possible example in section 5.5.

4. ACTING IN MOTION

Direct manual interaction is the standard input for computing devices. Examples are keyboards for text input, hardware buttons for special functions, as well as touchscreens with arbitrary virtual widgets. Although hand-held touchscreens in general allow for intuitive usability, they are not always appropriate for players on the move.

Large-scale motion is an inherent interaction mode for location-based games. Hence the computing system already has to support location sensing and processing. Beyond position a fruitful context parameter to measure is orientation. Digital compasses are available either as separate small sensors or integrated in mobile phones as the Nokia 5140.

Gestures based on small-scale body motion can be used in different ways for input. (1) the device can be moved systematically, e. g. tilting or shaking the device can be sensed by accelerometers and used to detect gestures. Stegmann et al. discuss in [12] how they successfully apply motion control and other modalities in the course of the MediaScout project of the Deutsche Telekom Laboratories to enhance media access. The implementation of some basic gestures for using a mobile phone as virtual music instrument is described in [5]. (2) mobile devices could be used as pointing devices in the real environment as Simon et al. discuss in [11]. Based on such reality pointers gestures could also be perceived in the context of the ambience.

Audio inputs can be used in terms of voice control (active or interactive) or noise sensors. Speaker independent speech recognition is feasible when the vocabulary is small enough. Audio commands can be recognized on the handset, but due to the limited resources of mobile devices, true voice recognition requires redirecting a powerful server as with the VLingo system. For most games a limited set of commands is already very helpful and allows even for steering as for example the Vocal Joystick project shows, c. f. [6].

5. SOME TYPICAL GAME SITUATIONS

In the following we present examples of reoccurring pertinent game situations that can be supported by certain modality mixes. Some of these scenarios have already been implemented and tested, others are planned and some not realistic yet and need further research. The scenarios are analyzed with respect to determining contexts.

Beside these particular conditions, the game itself is a very special context that influences the adequateness greatly. This holds in particular when mobile phones are used and

[4]The heat which some mobile devices, especially first-generation UMPCs, produce when running is currently not directly controllable with an API.

[5]http://www.painstation.de

the user is accustomed to communication-specific interaction modes, as the *in game communication* situation illustrates.

5.1 On the Move

As we pointed out in section 3, motion limits visual perception, thus an adaptation towards auditive and tactile interaction is meaningful. Headsets and vibration units are widespread but typically used for passive phone application only which is achieved with one button on the headset.

For game interaction normally a larger set of commands should be instantly accessible. This demands further interaction than passive phoning does. Accelerometers are already being used for the selection of commands, e. g. the UTI project at T-Labs where tilting a remote control is used to browse through a menu hierarchy (c. f. [12]). This application relies on visual feedback. In our scenario auditive or tactile feedback would be more appropriate. A limited number of commands can be used with tactile feedback only as illustrated in figure 1. For complex command sets like menu structures, audible feedback for the selection of certain commands is an option.

Figure 1: Tilt the device in different directions to select commands or browse menus.

In the case of *Scotland Yard to go* the set of commands that are likely to be demanded by a detective on the move contains for example *establishing a conference call to all detectives*, *communicating the fact that Mr. X is very close*, *asking the other detectives to help and come close.*

5.2 Approaching a Suspense Climax

In games with multiple parallel threads of action often unforeseen climacteric situations emerge. Multiplayer games like *Scotland Yard to go* fall into this category and the relative location of detective and Mr. X often leads to critical situation where his arrest hangs by a hair. Two urgent demands occur now conjointly: movability and awareness of the mutual opponents. Conventional user interfaces do not suit these needs in common. A fleeing Mr. X is neither able to observe a display nor can he trigger a rescuing action by typing a button (e. g. throw a virtual "fog bomb" to be out of the detective's displays for some minutes).

We aim at a solution by modulating the frequency or vibrations or acoustic signals. This has already been applied successfully in the Ambient Life project to continuously communicate the state of a phone by life-like signals, c. f. [7]. As input triggering an important action gestures performed with the device as a whole can be recognized through accelerometers. This sort of interaction seems to be much more appropriate and easy to manage in the presence of a hectic atmosphere and appears much more intuitive.

5.3 Hearkening and Tiptoeing

When hiding is an issue, hearkening and tiptoeing are natural behaviors. In a computer supported mobile game this could be integrated in two different ways: simulating the situation and supporting the real situation.

Let's first look at hearkening and tiptoeing as a virtual situation. We can simulate the situation even while players are in reality too far away. The pace of the hiding player can be sensed by accelerometers. In case he moves too fast, the noise of his paces can be simulated and communicated to the other players. For many smartphones software that tracks the pace is readily available (c. f. [9] and [4]). Second, if the situation occurs in reality the interaction modality should be adapted. When Mr. X is tiptoeing because some detectives are very close to him he is careful not to make a noise. In this situation audio would be a misplaced media – even if he was running shortly before and therefore auditive interaction was appropriate.

5.4 In-Game Communication

Oral communication is powerful and important. In case of *Scotland Yard to go* the detectives cooperate intensively, therefore conference calls are vital for the game experience. In some situations, Mr. X can apply wiretapping to be informed about the most recent theories and plans of the detectives. This is crucial for a rich game experience, since one of Mr. X's aims is to lead the detectives up the garden path and obtain himself the possibility to run away unnoticed.

We think about adapting the typical phone user interaction in these cases by (1) using a game-specific ring tone or vibration pattern; (2) in-game-calls can be automatically accepted without pressing a button.

5.5 Clues and Disturbing Puzzles

In adventure-like location-based games, clues are regular means to lead the players towards their goal. In motion-intensive games as for example in hide-and-seek games it could be exciting to combine the discovery of the helpful clue with a disturbance of the player's equilibrioception. In such a case, the embedding into the right situation is crucial to gain effective game experience. Assumed a player needs to solve a puzzle to get a hint for Mr. X's current hiding position. Figure 2 illustrates[6] how such a puzzle could look like. Such a puzzle could be placed best immediately before the player has to move on chasing Mr. X.

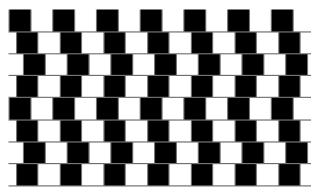

Figure 2: Count the straight lines to see Mr. X.

5.6 Surrounded by the Game

Often the relative position of a player to others or to game artifacts is crucial. In *Scotland Yard to go*, Mr. X is often

[6]Picture in figure 2 taken from www.eyetricks.com.

surrounded by detectives. It would certainly be a great experience to hear their voices or steps in the according directions and an intensity representing their distance.

Such a scenario is at least partially realizable. First of all beside the position of all involved players the orientation of Mr. X could be tracked by a digital compass as integrated for example in the Nokia 5140. The harder part is the sound output. The technique of in-ear recording with a dummy head is established since decades. Perception requires only usual earphones and recording is also easy as long as the relative locations of output and input are static as typically the case when visiting a concert or stage play.

Such a static setting reduces the applicability in location-based games drastically. Nevertheless, for some games situations static orientation of a player and its relative location to other players and artifacts can be preplanned in advance or do not have to be represented exactly but can be abstracted to a rather symbolic position. For example, when Mr. X is surrounded by detectives their distance can be communicated adequately by prefabricated surround sound. We expect the drawback of the inaccurate position representation to be negligible for the game experience.

Furthermore, we suggest to expend research efforts in order to render prerecorded sounds and life audio streams according to dynamic positions. The SWAN project already has developed a prototype using acoustic signals for auditive navigation (c. f. [13]). If this would be technically feasible for voice or music and in particular for live streams, even low sound quality would create a stunning experience and allow for completely new game design and interaction.

6. CONTEXT AND SITUATION IN GAMING

Bichard and Waern discuss the design of game activities against three main styles of games: *gamist* games that rely mostly on quests and puzzles, *simulationist* games where the player immerses with his own playful character into an evolving game story, and *dramatist* games that prescribe most of a story and let the player "listen" to the game. In [1] they present the dramatist game "Interference" which could be characterized as a pervasive adventure game.

If we follow this classification it seems reasonable that for gamist and dramatist games many situations can be predesigned in a screen-play. Those scenes could be supported by preplanned specific user interaction modalities. Such dramatist games often require a holistic design approach that resemble more drama authoring than technical event flow design. However, for scavenger hunt games and alike there are already some editors available (c. f. GPSMission[7] by Orbster or MediaScape[8]).

Simulationist games such as *Scotland Yard to go* do not predefine the story in such detail. To the contrary, some characteristic situations evolve during the play. Hence possible adaptations of the user interaction modes can only rely on situation detection at runtime. In the course of the project Context-Sensitive Intelligence we developed a prototypical architecture for such runtime adaptations (c. f. [10]).

7. CONCLUSION AND OUTLOOK

This paper argues that multimodal and intermodal interaction alone cannot cope with location-based games in particular when they are motion intensive. Awareness about the player's context and gaming situation is a key to select the appropriate interaction mode. We discuss how sensing and acting relates to the player's context and show examples for candidates of game design patterns that can be applied in specific situations. We conclude that context-sensitivity should be an integral part of game design methodology.

We developed this paper based on our experiences in designing *Scotland Yard to go*, a pervasive hide-and-seek game. Further we took a look at different game families as pervasive adventures, explorative games and rallies and defined the category of motion intensive games to focus on the characteristic that makes intuitive usability so difficult to achieve. Our suggestion is to apply multimodality in innovative ways and parallel it with context-sensitivity. This can be done in a predefined manner at game design time or in a reflective way at runtime or a mix of both.

8. REFERENCES

[1] J.-P. Bichard and A. Waern. Pervasive play, immersion and story: designing interference. In *DIMEA, Athens, Greece*. ACM, September 2008.

[2] S. Brewster. *Nonspeech auditory output. In Sears, A. and Jacko, J. (Eds.) The Human ComputerInteraction Handbook*, chapter 13, pages 247–264. Lawrence Erlbaum Associates, USA, 2nd edition, 2008.

[3] S. Brewster and L. Brown. Tactons: structured tactile messages for non-visual information display. In *Proceedings of Australasian User Interface Conference*, pages 15–23. Australian Computer Society, 2004.

[4] Edovia Inc. *Steps*. www.edovia.com/steps/, July 2008.

[5] G. Essl and M. Rohs. Shamus - a sensor-based integrated mobile phone instrument. In *Proceedings of the International Computer Music Conference*, 2007.

[6] S. Harada, J. A. Landay, J. Malkin, X. Li, and J. Bilmes. The vocal joystick: Evaluation of voice-based cursor control techniques. In *Proc. of ASSETS*, Portland, Oregon, USA, October 2006.

[7] F. Hemmert. Ambient life: Calm and excited pulsation as a means of life-like permanent tactile status display in mobile phones. In *Proceedings of the Design & Emotion Conference*, Hong Kong, 2008.

[8] F. Hemmert, A. Knörig, G. Joost, and R. Wettach. Dynamic knobs: Shape change as a means of interaction on a mobile phone. In *CHI 2008 Proceedings*. ACM, April 2008.

[9] O. Kirkeby and M. Kähäri. *Activity Monitor*. Nokia Research Center, 2008.

[10] H. Mügge, T. Rho, and A. B. Cremers. Integrating aspect-orientation and structural annotations to support adaptive middleware. In *Proceedings of the 1st workshop on Middleware-application interaction: in conjunction with Euro-Sys 2007*. ACM, 2007.

[11] R. Simon, H. Kunczier, and H. Anegg. Towards orientation-aware location based mobile services. In *3rd Symposium on LBS and TeleCartography*, 2005.

[12] J. Stegmann, K. Henke, and R. Kirchherr. Multimodal interaction for access to media content. October 2008.

[13] B. N. Walker and J. Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors*, 48(2):265–278, 2006.

[7] gpsmission.com

[8] www.mscapers.com