Nonvisual, distal tracking of mobile remote agents in geosocial interaction

Steven Strachan and Roderick Murray-Smith

[1] Orange Labs - France Telecom
28 Chemin du Vieux Chne, 38240 Meylan, France
steven.strachan@gmail.com,
[2] University of Glasgow,
Department of Computing Science,
Glasgow, Scotland, UK.
rod@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~rod/

Abstract. With the recent introduction of mass-market mobile phones with location, bearing and acceleration sensing, we are on the cusp of significant progress in location-based interaction, and highly interactive mobile social networking. We propose that such systems must work when subject to typical uncertainties in the sensed or inferred context, such as user location, bearing and motion. In order to examine the feasibility of such a system we describe an experiment with an eyes-free, mobile implementation which allows users to find a target user, engage with them by pointing and tilting actions, then have their attention directed to a specific target. Although weaknesses in the design of the tilt–distance mapping were indicated, encouragingly, users were able to track the target, and engage with the other agent.

1 Introduction

With the recent introduction of mass-market mobile phones such as the Nokia *6210 Navigator* with location, compass bearing and acceleration sensing, we are on the cusp of significant potential progress in location-based interaction, and mobile social networking (Fröhlich *et al.* 2008). Currently this sensor group is primarily marketed for pedestrian navigation (Strachan *et al.* 2007, Jones *et al.* 2008), but it has the potential to be used for a much more exciting range of interaction styles. Strachan and Murray-Smith (2009) have described bearing-based interaction with content and services, and in linked work Robinson *et al.* (2008) describe its use for GeoBlogging, or 'Googling the real world', as one of their participants observed. It is also an obvious step to couple this with social networking applications, where users can probe and point at and engage with nearby friends (Strachan and Murray-Smith 2008). The richness of the sensing, and the context-sensitivity and person-specific nature of such communications suggest that designers should beware of implementing overly prescriptive mechanisms for allowing individuals to interact in such systems. We argue in this paper that representations which display the uncertainty in location, bearing and inferred context are necessary for the success of such systems, and that this allows performance to grow

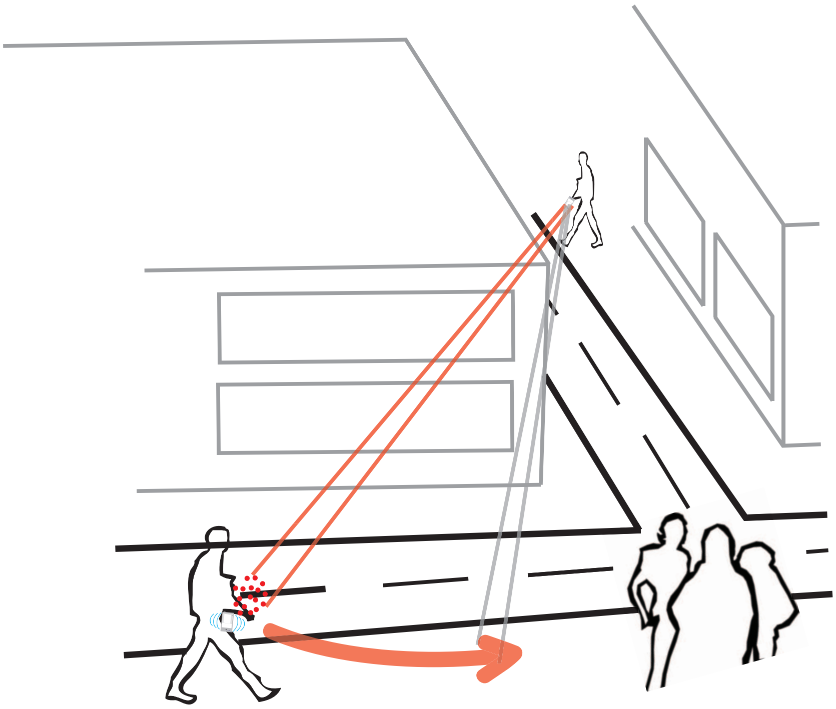


Fig. 1. A user interacts in the virtual environment with a friend, attempting to guide her to the group of friends already assembled. The interaction is primarily physical and non-visual, pointing with the phone with audio and haptic feedback as a function of the movements of the individuals.

over time as new sensing technologies and better context inference mechanisms are developed. This assumption needs to be tested empirically, and the research community needs to evolve methods for reliably comparing competing designs for such geosocial tasks. This paper introduces and describes a system, with an initial user study, which examines the interaction between a human using embodied bearing-based interaction with feedback generated by a simulated model of a human, testing whether it is in fact possible for users to track others in the virtual environment with realistic sensing conditions.

2 Mobile Spatial Interaction

Mobile Spatial Interaction (MSI) is a form of interaction that enables users to interact with a hybrid physical/virtual environment using their mobile device. Users are given the ability to interact, in an eyes-free manner, with the virtual environment by using their mobile device to focus on an object in the real world or scan the space around them to discover virtual objects using pointing motions as illustrated in figure 2. The direction in which the user is pointing is taken from a compass heading estimated using magnetometers with accelerometers used to compensate for tilt.

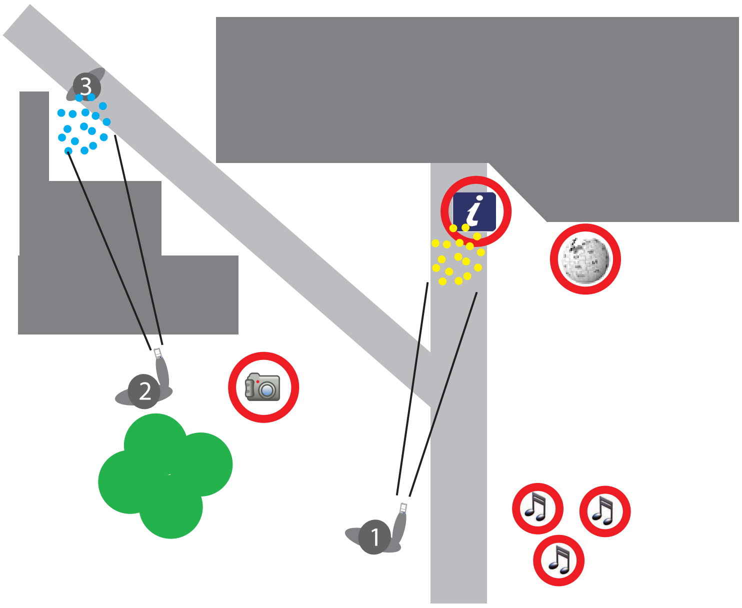


Fig. 2. User 1 is interacting with information in their combined virtual/physical environment while user 2 is interacting with user 3 out of their line of sight.

2.1 Virtual Environments

A fluid and unrestricted collaboration between two or more people connected remotely via a computer has long been a goal in the fields of virtual and augmented reality. Collaborative Virtual Environments (CVEs) (Benford *et al.* 2001) enable a sense of shared space and physical presence in the virtual world. The increasing power and ubiquity of continually connected, location-aware and continuously sensing mobile devices has now enabled us to generalise down to the mobile realm with the development a Mobile Collaborative Virtual Environment (MCVE). We present in this paper an MCVE that enables the connection of two or more mobile devices to create a hybrid 'eyes-free' physical/virtual world in which users may interact using their mobile device as a probe for objects or for other users located in the virtual environment, while all the time receiving audio and vibrotactile feedback dependent on the nature of their probing. A key aspect to the success of this kind of interaction is the provision of a sense of embodiment or presence in the virtual environment. Greenhalgh and Benford (1995) tackle this with the DIVE and MASSIVE systems by providing a number of graphical representations of embodied participants. The major advantage that these systems have is that they are highly visual and significant emphasis is placed on the provision of visual feedback to the user. Much of the work conducted on eyes-free systems is for the visually impaired. Magnusson and Rassmus-Gröhn (2005) describe a haptic-audio system designed to guide visually impaired users through a traffic environment for exploring and learning a route. They find that most users were able to navigate a fairly complex virtual model with little trouble. Crossan and Brewster (2006) describe a system for two-handed navigation in a haptic virtual environment designed for visually impaired users finding that participants were able to traverse a maze using haptic feedback alone.

Social Cognition in the Virtual Environment One of the major functions of social cognition in humans is to allow the creation of a shared world in which interaction can take place. Communication between two or more people is greatly enhanced by the the adoption of a shared vocabulary that enables us to share goals, so that we may then engage in joint activity. For successful interactions to take place it is necessary that the interactors achieve the same perception of the world, referred to as 'common ground' (Clark 1996). While this is not easy to achieve in a completely abstracted virtual environment, in a hybrid virtual/physical environment this is slightly easier to achieve. Since an MCVE is located in the real world the user is not completely immersed in the virtual world, they have access to real-world visual cues and so some of this natural intuition regarding the interaction with the physical world may be transferred into the virtual world. The augmentation of real world objects is one approach to providing a more effective interaction with the virtual world. Espinoza *et al.* (2001) describe their GeoNotes system that allows users to leave virtual messages linked to specific geographical positions. They strive here to socially enhance digital space by blurring the boundary between physical and digital space. But still little has been achieved in terms of active interaction or collaboration between two or more users in this kind of environment and it remains a challenge. The starting point for this kind of active and collaborative interaction is to align the focus of our attention in the environment, typically achieved in real life by pointing at an object or watching bodily movements that can give us some idea about the intentions of the other person (Frith and Frith 2006). Our bodies are used to provide continuous and fine-grained social cognitive signals about our psychological state, our presence, activity and our attention via gestures, facial expressions or general body posture. It is important then that this kind of information is not lost completely in the virtual environment.

3 Description of Experimental System

The system described here builds on earlier work (Strachan and Murray-Smith 2009), such that the location is provided by the GPS sensor and bearing by a combination of magnetometer and accelerometer. This means that we can point in any direction with the device at any orientation. The orientation (pitch and roll) of the device is estimated using the x, y and z components of the acceleration with respect to the gravity vector, and the pitch angle is used to control how far into the distance the user is pointing with the 'virtual probe', illustrated in figure 2. If the user wishes to look further ahead into the space in front they tilt the device forward. If they wish to bring their probe back to their current position they tilt the device back again, effectively pulling the device to their chest as illustrated in figure 3. The actual distance looked ahead is linearly mapped to the pitch angle of the device with a 90° pitch mapping to a 0m look-ahead and 0° pitch mapping to a 30m look-ahead. A user now has the ability to obtain information about the space around them by listening and feeling for impact events (via audio and haptic feedback from the device), when their 'virtual probe' effectively collides with objects in the virtual environment.

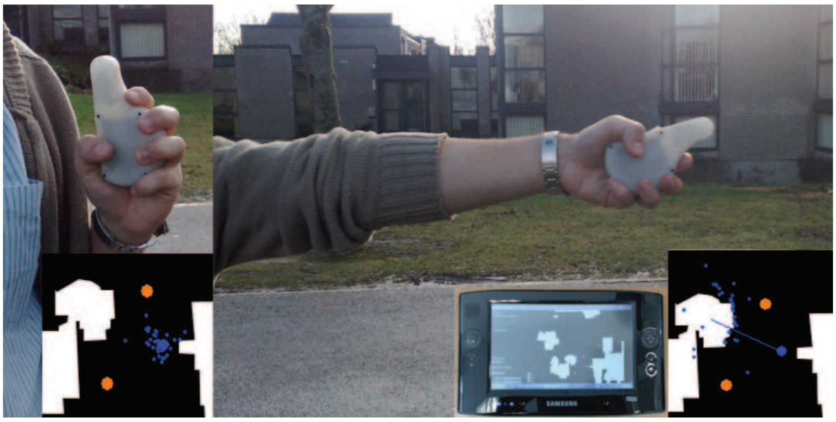


Fig. 3. Varying the orientation of the device alters the distance at which the user is probing.

One further feature of this system is that it allows an agent to interact remotely with other agents in the virtual world by pointing and interacting with their mobile devices as illustrated in figure 2. For the purposes of this study interaction is limited to discrete probing of the other agent's device but it is possible in future iterations to expand this to include probing of the other user's movements giving more information about the specific areas of interest and intentions of the other user.

3.1 Uncertainty and the virtual probe

With this kind of continuously sensing system, uncertainty becomes an important factor to consider. We are subject to uncertainty in our GPS position estimate, uncertainty in the signals from our sensors, uncertainty in the user's motor actions, uncertainty in the intentions of the user and uncertainty in the system's belief about the user's intention that is fedback to the user. If these uncertainties are not considered and dealt with appropriately they have the potential to render a system of this kind unusable.

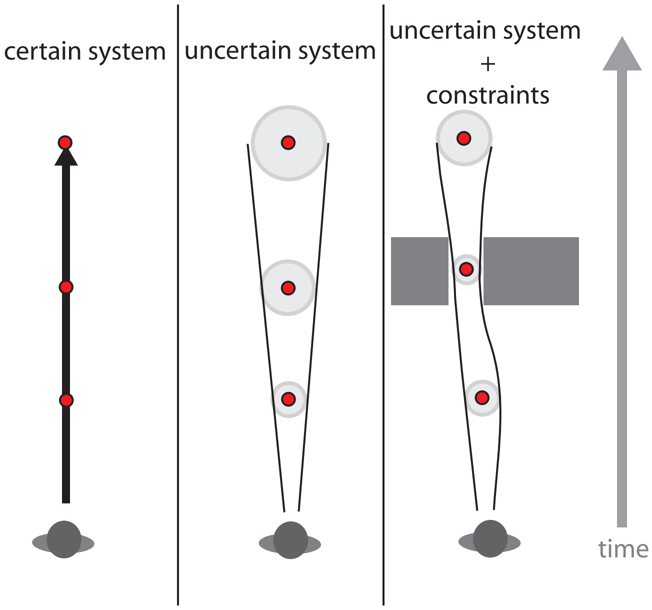


Fig. 4. Left: As we move forward in time (from bottom to top), if there was no uncertainty in our position or direction of travel, we would know exactly where we were going to be in a specified amount of time. Center: When we include uncertainty in the estimates of our position and direction we become less and less sure of our future position as we move forward in time. Right: If we include constraints, such as roads or gaps between buildings, as well as uncertainty, we can reduce the uncertainty in the constrained areas.

The system described in this paper explicitly uses the display of uncertainty to assist a user's navigation of their current location or context by acknowledging all of the uncertainties mentioned above. A straightforward propagation of our user's position through the space in front would lead to a certain point at some specified time horizon (figure 4: Left). This does not model the user's potential future positions effectively and is likely to lead to confusion from the user when uncertainty in the position estimate, for example, suddenly increases and the seemingly certain estimate is now wrong. For example, if a user is moving North at a pace of 5m/s, in 1 minute in a completely certain system he will be exactly 300m North of his current position. But if we take into account any uncertainties, in his initial position, in his heading, in his intentions and any constraints that lie in his path (such as buildings or busy roads) we are much less certain about where he will be in 1 minute (figure 4: Center, Right). In this case we know he will be roughly 300m away but there will be some distribution or uncertainty around that position which increases as we look further ahead in time. By projecting possible user paths into the future from some location along a given heading using Monte Carlo sampling, with uncertainty injected into both the location and heading estimates and constraints in the environment included we are presented with a distribution (represented by Monte Carlo sampled particles as in figure 2), which represent the most likely position of the user in a specified amount of time in a specific context as illustrated in figure 4.

With most currently available location-aware navigation software a single best estimate of position is made but this gives the user unreasonable confidence in the accuracy of their system and prevents the interactor from choosing an optimal strategy for dealing with the true state of the system. It has been shown that unrealistically precise feedback in the presence makes smooth, stable and accurate control difficult (Williamson *et al.* 2006) and (Strachan *et al.* 2007) and it is necessary to acknowledge that the kind of real world uncertainties we are exposed to when designing this kind of interaction place considerable limitations on the overall interaction (Strachan and Murray-Smith 2009).

A convenient side effect of this Monte Carlo propagation is that the future predicted positions can be used as a virtual probe for our virtual environment with the current future time horizon controlled with the pitch of the device and the direction taken from the compass heading as described above. By providing the user with the ability to move this cloud anywhere in the virtual environment using this functionality and feeding back any interaction between the cloud and objects of interest using audio and haptic feedback, we enable a functional and embodied style of interaction.

3.2 Hardware

The current system runs on a Samsung Q1 Ultra Mobile PC with a Bluetooth connection to the WayStane (Murray-Smith *et al.* 2008) inertial sensing pack as shown in figure 5. This device contains the magnetometers, accelerometers and vibration devices that we require for this kind of interaction. The device also contains gyroscopes and capacitive sensing and is an adaptation of the SHAKE (Sensing Hardware Accessory for Kinaesthetic Expression) inertial sensing device.

4 Experiment

An experiment was conducted in order to quantify the users' ability to track a moving target, displaying varying behaviour, in a virtual environment using the described system with only audio and tactile (i.e. eyes-free) feedback.

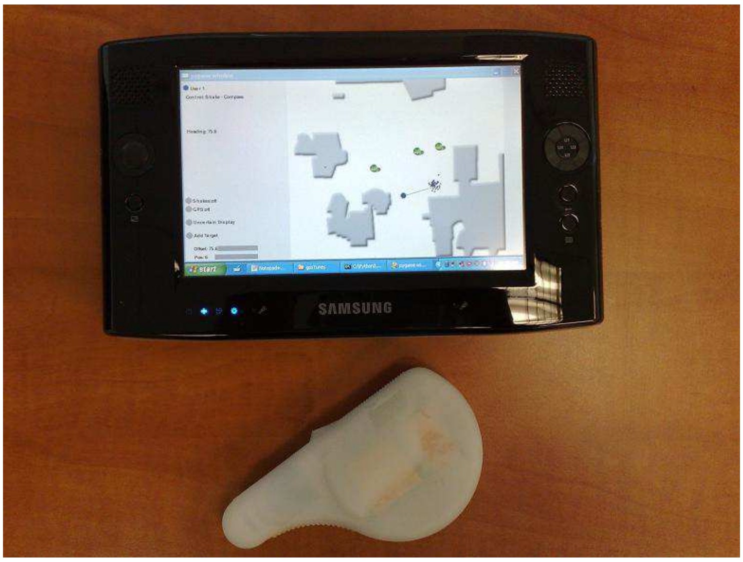


Fig. 5. A Samsung Q1 UMPC with a Bluetooth connection the WayStane inertial sensing device

Participants were first given an introduction to the system and a visual example before being allowed to practice using the functionality of the device. They were then blindfolded and placed in a specific spot within a large room which represented the center of a circle within the virtual environment, corresponding to a circle in the real world with an approximately 65m diameter. They did not move from this spot. They were then given the task of tracking a simulated human agent using the functionality of the inertial device. A screen shot is shown in Figure 7.

The modelled human agent displayed three kinds of behaviour, illustrated in Figure 6. Here we create a basic model of possible walking behaviour. In the first phase the agent is given random walk dynamics. From some initial position the agent moves by a distance drawn from a Gaussian distribution with a mean offset of 1.1m and a standard deviation of 0.5m. This distribution was observed from a GPS trace of one person walking at a normal pace outdoors in good GPS visibility. The agent was given a heading update drawn from a Gaussian distribution with mean 0 and standard deviation of 3.07°. This value was chosen to allow the agent the freedom to change direction without any implausible jumps in the heading which would lead to an erratic and unnatural walking pattern. The agent was also constrained to stay within the virtual circle.

The experiment participant was first helped to locate the agent, which is then tracked using the virtual probe for a short time until a switch to a second *attention check* mode is activated. This switch corresponded to the point where the user had sufficient contact (i.e. when the energy of impacted particles was above some threshold) with the agent for approximately 15s in total. The *attention check* part of the agent's behaviour was designed to detect if the participant could explicitly pick up any unusual systematic hops in order to prove that the participant had really discovered the target. Such a hop might be typical of a gesture which a user might generate to indicate recognition of

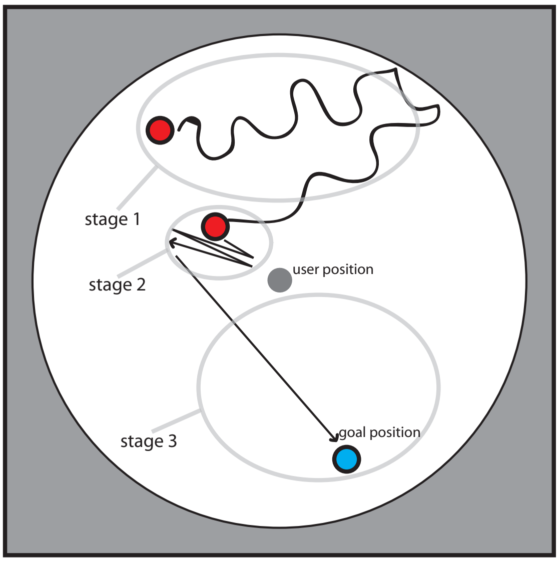


Fig. 6. The three stages of modelled human behaviour. Stage 1 shows the “random walk“ phase where the agent randomly moves around the circular are until it is detected by the user. Stage 2 shows the “attention check“ phase where the agent consecutively jumps from left to right in order to check if the user follows these jumps. Stage 3 shows the “goal-directed“ phase where the agent moves straight towards a target point as long as the user is tracking the movement.

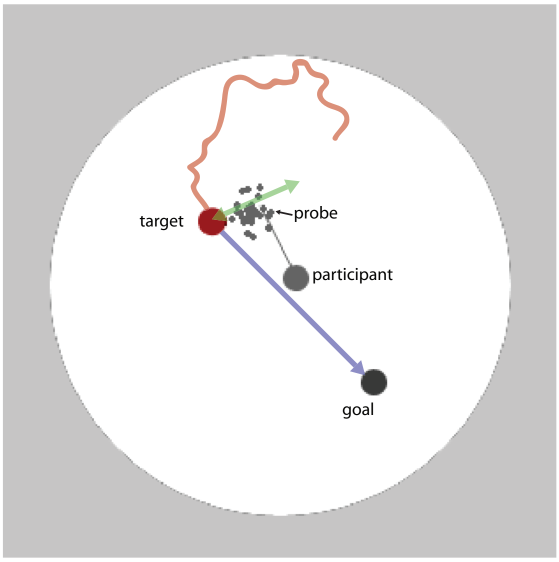


Fig. 7. Screen shot of user tracking an agent using their virtual probe. The red, green and blue superimposed lines indicate the random walk, attention detect and goal-directed stages respectively.

engagement, and which through imitation might propagate through a geosocial subculture. Immediately after the switch to this phase the agent consecutively jumps to the left, is detected by the participant, then to the right and is detected again for a total of 5 hops. Each hop corresponded to approximately 2.5m in the real world. This hop was designed to be large enough to be distinguishable from the movement in the first phase of agent behaviour. When the five hops were all successfully detected the agent switches to phase three, the goal-directed phase.

The *goal-directed* phase involves the movement directly to a particular place within the virtual environment. The agent is given the precise bearing of the target area and moves directly to that point only if the participant has sufficient contact with the agent. This is intended to test whether people could have their attention guided to a particular location. If the participant loses contact, the agent remains static until it is found once more. When the participant has tracked the agent to this target the experiment is complete. Participants who were performing particularly well after reaching the first goal were given up to two more goals to reach.

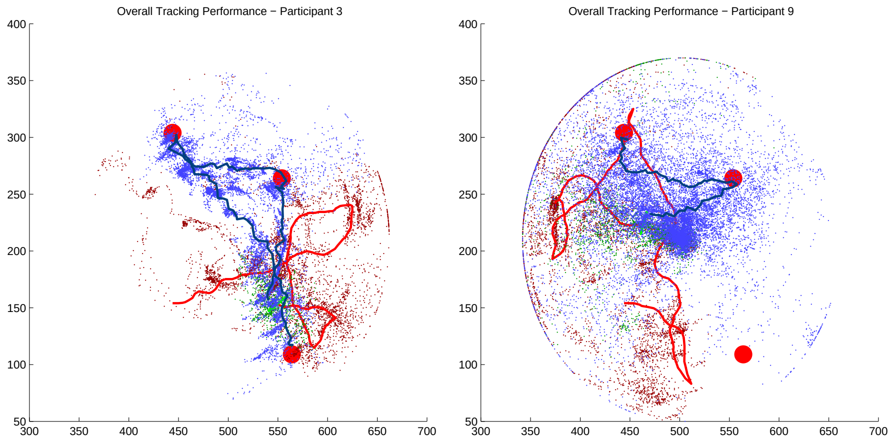
It is likely that if the experiment was conducted outdoors, it would lead to a more subjective analysis of user performance since we could not be aware of exact uncertainty present in the system. To enable a more objective comparison of user performances in a strictly controlled environment, we added uncertainty to the positions simulated, to create as realistic a situation as possible, testing whether this kind of interaction would be feasible in a realistic mobile spatial interaction system based on current commercially available technology. The uncertainty distribution of the participant's probe, as described in the uncertainty section was given a value in the heading estimate of 2.29° and a value of approximately 1.5m in the position estimate. Noise was also added to the participants position in order to simulate the effects of being in the real world. A value of approximately 2m was also added to the participants' virtual GPS position at 1Hz, which produced a hopping effect on the participant's position. Although this is not strictly the effect we would observe from a real GPS position estimate, which is likely to shift much more gradually, and systematically over time, the method used was considered more consistent over all of the participants and was unlikely to lead to large periods of very difficult interaction for some users and not for others. 13 participants in total took part in the experiment ranging from 20-60 yrs of age with 10 males and 3 females.

4.1 Results

A wide range of performance was observed throughout the trial. 9 participants finished the whole task with all 3 goals, 4 achieved at least one target in the goal directed stage.

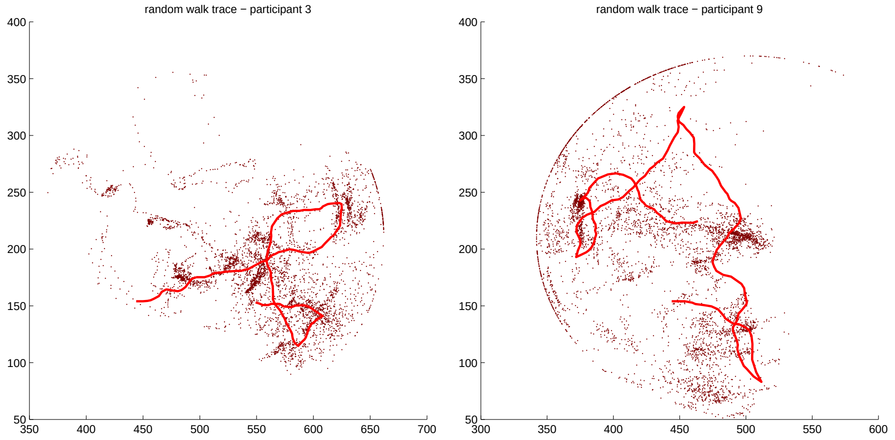
Figure 8 shows the difference observed over the whole run for one participant who performed very well and one who performed less well. It is clear to see that participant 3 stuck closely to the target path, shown in more detail in figures 9(a) and 10(a) whereas participant 9 performed a much longer search and covered a much larger area of the environment.

Figures 11(a) and 11(b) show the normalised mean squared error (i.e. the difference between the agent's position and the user's search position) in the tracking performance of each participant. It is clear that the participants who performed better overall, i.e. 2,



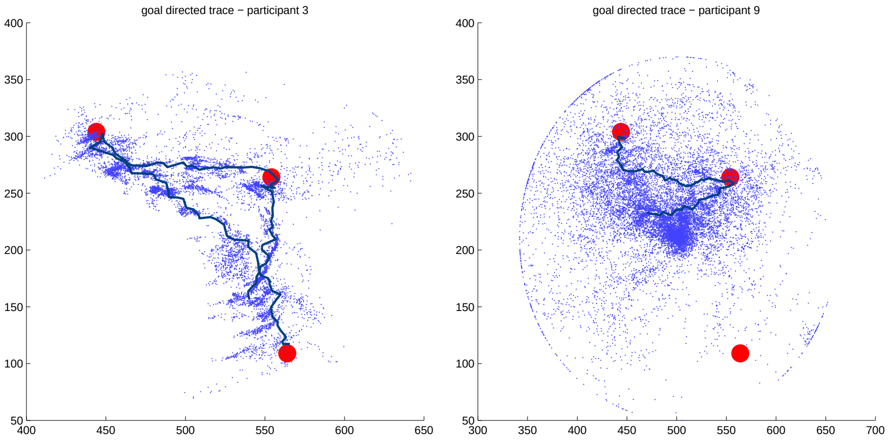
(a) (b)

Fig. 8. (a) The full tracking run for participant 3 who showed a good overall performance. (b) The full tracking run for participant 9 who showed a bad overall performance. Red denotes the random walk phase. Green denotes the attention detection phase and blue denotes the goal-directed phase. The small dots indicate the area scanned by the user in each phase and the solid line indicates the position of the agent in each phase of the experiment. The large red circles indicate the locations of the goal areas.



(a) (b)

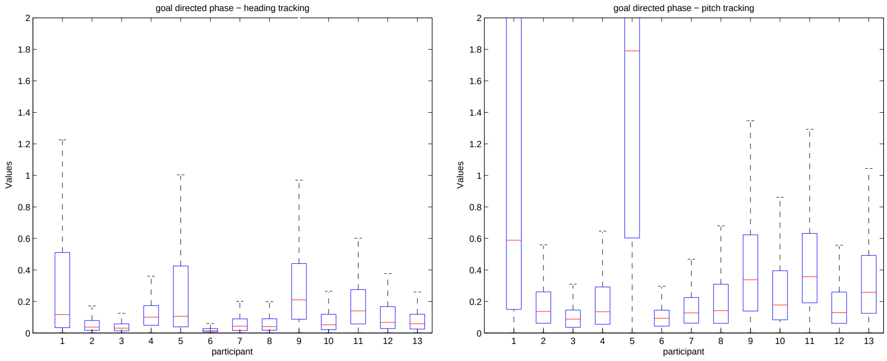
Fig. 9. Tracking performance for the random walk phase only, for a good (a) and bad (b) performance.



(a) (b)

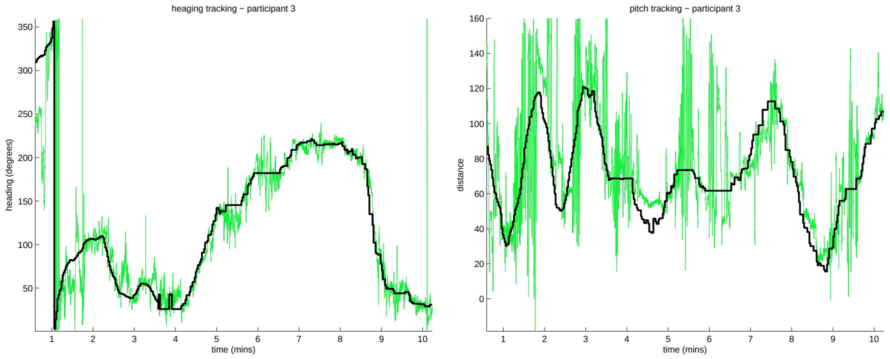
Fig. 10. Tracking performance for the goal directed phase only. One participant completes all 3 targets whereas the other fails to complete all 3.

3, 6 and 7 have a much lower mean error meaning that they tracked the agent much more closely overall. The standard deviation, indicated by the size of the boxes in Figures 11(a) and 11(b) is also lower for these participants, again indicating that there was much less variation over the whole run. Participants with higher mean values and larger standard deviations are those who performed the least well. When examining



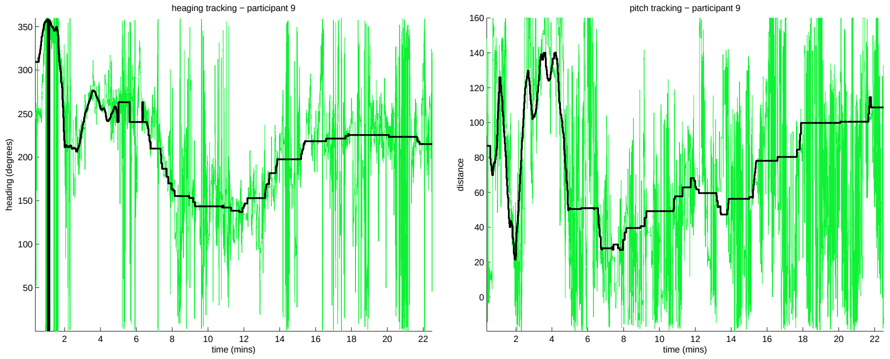
(a) (b)

Fig. 11. (a) The mean squared error for all participants in the heading tracking task. (b) The mean squared error for all participants in the tilt tracking task.



(a) (b)

Fig. 12. (a) The participant tracks the heading of the agent very closely but struggles to track as effectively with the pitch (b). The green line represents the participants current heading and the black line represents the bearing of the agent.



(a) (b)

Fig. 13. (a) The participant performs badly in both the heading and pitch tracking tasks (b). The green line represents the participants current look-ahead distance (mapped to the device pitch) and the black line represents the bearing of the agent.

the performance of participants for the tracking task in both the heading and pitch cases separately it is clear to see from figures 12 and 13 that users generally performed better when tracking the bearing of the agent than they did while tracking the distance of the agent. Figures 12(a) and 13(a) show the heading tracking performances for participants who performed well and poorly respectively and figures 12(b) and 13(b) show the distance tracking performances for the same participants, which shows a clear difference in performance. This backs up comments made by a number of the participants who claimed that they had more problem tracking the distance (back-forward movements) than the bearing of the agent and this is probably due to the fact that appropriate feedback was provided for the bearing tracking task (left-right audio pan) but not the distance tracking task.

5 Discussion and Conclusions

This paper introduced a new form of embodied geosocial interaction using current technology and demonstrates the feasibility of this new form of interaction via a user study that simulates both realistic environmental conditions and human behaviour. The system presented explicitly displays uncertainty in location, bearing and inferred context.

With the experiment it was found that users, with little experience or practice with this kind of system, were able to track a modeled human agent through three different types of behaviour. A number of issues were highlighted. By far the most common problem commented upon by the participants was the difficulty in tracking an agent which was moving towards or away from them. No participants had any significant problem with the heading tracking metaphor since this was a completely natural pointing technique with which we are all familiar. But since audio feedback was only provided for actual contact with the agent and slight movements to the left and right (represented by audio panning) there was great difficulty in the perception of the forwards and backwards movement of the agent. It was clear that different participants were perceiving the behaviour of the agent in different ways even though the agent was given identical behaviour for each run. Some users commented that sometimes the agent was moving around very fast and other times it was a normal speed when in reality the agent was simply further away when it was perceived as being slower and closer when it was perceived as being faster indicating that these participants had an issue with the the metaphor used for distance scanning. An obvious remedy to this problem would be the provision of explicit feedback for the towards and away motion, perhaps through another channel of communication. Another remedy could also be the provision of a more natural metaphor than the pitch-distance mapping metaphor we used here. This issue does highlight though, the sensitivity that any results in location-aware interaction will have to very detailed aspects of interaction design, whether input or feedback.

This initial feasibility study has demonstrated that this new form of interaction is possible but there exists great potential for the further development of richer interaction design in this context. When designing novel forms of interaction in a virtual environment application it is possible to use the theory of social cognition to examine a user's interaction with the system and any emerging low level human interactions. For example, might we see any signs of social cognition in this kind of system? Social cognitive signals are almost subconscious for humans in the real world but how does this translate into the virtual world? The processes of joint attention and imitation, for example, are fundamental to the investigation of early human behavioral patterns. The ability to infer intentions from overt behaviour in geosocial systems will allow users the potential to evolve new styles of communication, and to use the systems in ways the designers might not have anticipated. This will range from straightforward cases such as strangers detecting a common focus of attention, which allows them to begin interaction with each other, to subtle ongoing turn-taking and imitation between friends who know each other well, and who, given their mutual context awareness, can communicate significant amounts of information about their mood and intentions with subtle movements of their mobile phone.

The experiment in this paper used a simulated agent and a human user, in order to improve repeatability and for more control of the activity levels. It is important to note that in this experiment the user was using their hand to track only the position of the agent. Hand movements are more expressive and more rapidly changing than location, and so the generalisation of this work to mutual interactions between two user's rapidly changing hand movements will require careful consideration. This first step should act as the starting point for further work on joint attention in geosocial interaction. The next challenge is to expand this work to include multi-user interaction to validate the results of this paper with two human users, and to observe the detailed interactions that evolve as people engage and disengage from remote contact with each other.

Acknowledgements Both authors had positions with the Hamilton Institute, NUI Maynooth when this research was carried out. We are grateful for support from: SFI grant 00/RFP06/CMS052, EPSRC project EP/E042740/1, *Multimodal, Negotiated Interaction in Mobile Scenarios*, IST Programme of the European Commission, under PASCAL Network of Excellence, IST 2002-506778, and the EC-funded *OpenInterface* Project. Nokia provided a donation of funds and equipment.

References

Benford, S., C. Greenhalgh, T. Rodden and J. Pycock (2001). Collaborative virtual environments. *Commun. ACM* **44**(7), 79–85.

Clark, H. (1996). *Using Language*. Cambridge University Press.

Crossan, A. and S. Brewster (2006). Two-handed navigation in a haptic virtual environment. In: *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*. ACM. New York, NY, USA. pp. 676–681.

Espinoza, F., P. Persson, A. Sandin, H. Nyström, E. Cacciatore and M. Bylund (2001). Geonotes: Social and navigational aspects of location-based information systems. In: *UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing*. Springer-Verlag. London, UK. pp. 2–17.

Frith, C.D. and U. Frith (2006). How we predict what other people are going to do. *Brain Research, Multiple Perspectives on the Psychological and Neural Bases of Understanding Other People's Behavior* **1079**, 36–46.

Fröhlich, P., L. Baillie and R. Simon (2008). Realizing the vision of mobile spatial interaction. *interactions* **15**(1), 15–18.

Greenhalgh, C. and S. Benford (1995). Massive: a distributed virtual reality system incorporating spatial trading. *Distributed Computing Systems, 1995., Proceedings of the 15th International Conference on* pp. 27–34.

Jones, M., S. Jones, G. Bradley, N. Warren, D. Bainbridge and G. Holmes (2008). Ontrack: Dynamically adapting music playback to support navigation. *Personal and Ubiquitous Computing* **12**, 513–525.

Magnusson, C. and K. Rassmus-Gröhn (2005). A virtual traffic environment for people with visual impairment. *Visual Impairment Research* **7**(1), 1–12.

Murray-Smith, R., J. Williamson, T. Quaade and S. Hughes (2008). Stane: synthesized surfaces for tactile input. In: *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. ACM. New York, NY, USA. pp. 1299–1302.

Robinson, S., P. Eslambolchilar and M. Jones (2008). Point-to-GeoBlog: gestures and sensors to support user generated content creation. In: *Proceedings of the 10th international Conference on Human Computer interaction with Mobile Devices and Services (Amsterdam, The Netherlands, September 02 - 05, 2008). MobileHCI '08*. ACM, New York, NY. pp. 197–206.

Strachan, S. and R. Murray-Smith (2008). Geopoke: Rotational mechanical systems metaphor for embodied geosocial interaction. In: *NordiCHI '08: Proceedings of the fifth Nordic conference on Human-computer interaction*. ACM. New York, NY, USA. pp. 543–546.

Strachan, S. and R. Murray-Smith (2009). Bearing-based selection in mobile spatial interaction. *Personal and Ubiquitous Computing, Vol. 13, No. 4*.

Strachan, S., J. Williamson and R. Murray-Smith (2007). Show me the way to monte carlo: density-based trajectory navigation. In: *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. New York, NY, USA. pp. 1245–1248.

Williamson, J., S. Strachan and R. Murray-Smith (2006). It's a long way to Monte Carlo: probabilistic display in GPS navigation. In: *MobileHCI '06: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. ACM Press. New York, NY, USA. pp. 89–96.